(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 October 2002 (10.10.2002)

PCT

(10) International Publication Number
WO 02/079519 A1

(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number: PCT/US02/09752

(22) International Filing Date: 29 March 2002 (29.03.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/823,181 30 March 2001 (30.03.2001) US

(71) Applicant: **THE TRUSTEES OF COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK** [US/US]; West 116th Street and Broadway, New York, NY 10027 (US).

(72) Inventors: **JU, Jingyue**; 267 Marietta Street, Englewood Cliffs, NJ 07632 (US). **EDWARDS, John, Robert**; 485 Central Park West, Apt. 3C, New York, NY 10025 (US). **LI, Zengmin**; 647 West 172 Street, Apt. 41, New York, NY 10032 (US).

(74) Agent: **WHITE, John, P.**; Cooper & Dunahm LLP, 1185 Avenue of the Americas, New York, NY 10036 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/079519 A1

(54) Title: HIGH-FIDELITY DNA SEQUENCING USING SOLID PHASE CAPTURABLE DIDEOXYNUCLEOTIDES AND MASS SPECTROMETRY

(57) Abstract: This invention provides methods for sequencing DNA by detecting the identity of a nucleotide within a DNA sequencing fragment using mass spectrometry. The invention provides cleavable linkers for attaching a label to a dideoxynucleotide and provides labeled dideoxynucleotides. The invention also provides methods for increasing mass spectrometry resolution using linkers with different mass. The invention further provides systems for separating a labeled moiety from non-labeled components in one or more samples in solution.

HIGH-FIDELITY DNA SEQUENCING USING SOLID PHASE CAPTURABLE DIDEOXYNUCLEOTIDES AND MASS SPECTROMETRY

Background Of The Invention

Throughout this application, various publications are referenced in parentheses by author and year. Full citations for these references may be found at the end of the specification immediately preceding the claims. The disclosures of these publications in their entireties are hereby incorporated by reference into this application to more fully describe the state of the art to which this invention pertains.

The ability to sequence deoxyribonucleic acid (DNA) accurately and rapidly is revolutionizing biology and medicine. The confluence of the massive Human Genome Project is driving an exponential growth in the development of high throughput genetic analysis technologies. This rapid technological development involving chemistry, engineering, biology, and computer science makes it possible to move from studying single genes at a time to analyzing and comparing entire genomes.

With the completion of the first entire human genome sequence map, many areas in the genome that are highly polymorphic in both exons and introns will be known. The pharmacogenomics challenge is to comprehensively identify the genes and functional polymorphisms associated with the variability in drug

masked by the high fluorescence signal from excess dye-labeled primers or dye-labeled terminators, and are therefore difficult to identify.

Mass spectrometry is able to overcome the difficulties (GC compressions and heterozygote detections) typically encountered when using capillary sequencing techniques. However, it is unable to meet the read length and throughput requirements for large scale sequencing projects. In addition, poor resolution prevents the sequence determination of large DNA fragments. At the present time, the read lengths are insufficient for *de novo* DNA sequencing and the stringent clean sample requirements for using mass spectrometry for DNA sequencing are not entirely met by existing procedures. For this reason, most of the reported mass spectrometry applications have focused on single nucleotide polymorphism (SNP) detection. Several methods have been explored to this end. The most common approach is to extend a primer by a single nucleotide and detect what was added. Another technique developed by Tang et al. (1999) involves immobilizing DNA templates on a chip and again extending one base to determine a particular SNP. The same group has explored the analysis of restriction fragments to determine multiple SNPs at once (Chiu et al. 2000). Each of these techniques has been limited to analyzing only a few fragments at a time due to current limitations in mass spectra resolution. While these methods are sufficient for determining a SNP at a particular base, they require previous knowledge of the preceding sequence for

DNA sequencing fragments are not eliminated and are introduced to the mass spectrometer. False stops occur sequencing when a deoxynucleotide rather than a dideoxynucleotide terminates a sequencing fragment. It has been shown that false stops and primers which have dimerized can produce peaks in the mass spectra that can mask the actual results preventing accurate base identification (Roskey et al. 1996).

The "lock and key" functionality of biotin and streptavidin is often utilized in biological sample preparation as a way to remove undesired impurities (Langer et al. 1981). To date these methods have involved attaching the biotin moiety on the 5' end of the primer or the sequencing DNA template for capture by streptavidin coated magnetic beads (Tong and Smith 1992, 1993). When the samples are purified, false stops and primers that can interfere with the resulting sequencing data are not eliminated.

In addition, a further drawback of previous mass spectrometry sequencing methods was the requirement of four separate reactions, one for each dideoxynucleotide terminator analogous to the approach used in dye-labeled primer sequencing.

Ideally, for sequencing with MALDI-TOF mass spectrometry, one would like to establish a procedure that allows sequencing reactions to be performed in one tube to simplify sample preparation, to use cycle sequencing to increase the yield of the DNA sequencing fragments, and to have a method that only isolates pure DNA sequencing fragments free from

Biotinylated dideoxynucleotides and streptavidin coated magnetic beads can be used to generate high quality sequencing mass spectra of Sanger cycle sequencing DNA fragments on a MALDI-TOF mass spectrometer. The method disclosed here provides an efficient way to eliminate false stopped DNA fragments and excess primers and salts in one simple purification step, while still allowing the use of cycle sequencing to generate a high yield of sequencing fragments. Furthermore, it avoids the above-mentioned pitfalls of gel electrophoresis.

The subject application discloses that mass-tagged dideoxynucleotides which are coupled with biotin or photocleavable biotin can increase the mass separation of the DNA sequencing fragments on the mass spectra, giving better resolution than previously achievable.

Also, this application discloses a method for creating streptavidin-coated porous channels that can be used in light directed cleavage of the biotin-streptavidin complex. This is important as present commercially available streptavidin coated magnetic beads are inadequate for photocleavage purposes, in that they are opaque to ultraviolet light.

The system disclosed herein provides a high throughput and high fidelity DNA sequencing system for polymorphism and pharmacogenetics applications. Compared to gel electrophoresis sequencing, this system produces very high resolution of sequencing fragments and extremely fast separation in the time

Summary Of The Invention

This invention is directed to a method for sequencing DNA by detecting the identity of a dideoxynucleotide incorporated to the 3' end of a DNA sequencing fragment using mass spectrometry, which comprises:

(a) attaching a chemical moiety via a linker to a dideoxynucleotide to produce a labeled dideoxynucleotide;

(b) terminating a DNA sequencing reaction with the labeled dideoxynucleotide to generate a labeled DNA sequencing fragment, wherein the DNA sequencing fragment has a 3' end and the chemical moiety is attached via the linker to the 3' end of the DNA sequencing fragment;

(c) capturing the labeled DNA sequencing fragment on a surface coated with a compound that specifically interacts with the chemical moiety attached via the linker to the DNA sequencing fragment, thereby capturing the DNA sequencing fragment;

(d) washing the surface to remove any non-bound component;

(e) freeing the DNA sequencing fragment from the surface; and

(f) analyzing the DNA sequencing fragment using mass spectrometry so as to sequence the DNA.

This invention provides a method for sequencing DNA by detecting the identity of a plurality of dideoxynucleotides incorporated to the 3' end of

linker to a 5-position of cytosine or thymine or to a 7-position of adenine or guanine.

The invention provides a system for separating a chemical moiety from other components in a sample in solution, which comprises:

(a) a channel coated with a compound that specifically interacts with the chemical moiety, wherein the channel comprises a plurality of ends;

(b) a plurality of wells each suitable for holding the sample;

(c) a connection between each end of the channel and a well; and

(d) a means for moving the sample through the channel between wells.

The invention provides a method of increasing mass spectrometry resolution between different DNA sequencing fragments, which comprises attaching different linkers to different dideoxynucleotides used to terminate a DNA sequencing reaction and generate different DNA sequencing fragments, wherein the different linkers increase mass separation between the different DNA sequencing fragments, thereby increasing mass spectrometry resolution.

ddTTP) can be used with any of the illustrated linkers.

Figure 5: Synthesis scheme for mass tag linkers. For illustrative purposes, the linkers are labeled to correspond to the specific ddNTP with which they are shown coupled in Figures 4, 6, 8, 9 and 10. However, any of the three linkers can be used with any ddNTP.

Figure 6: The synthesis of ddATP-Linker-II-11-Biotin.

Figure 7: DNA sequencing products are purified by a streptavidin coated porous silica surface. Only the biotinylated fragments are captured. These fragments are then cleaved by ultraviolet irradiation (*hv*) to release the captured fragments, leaving the biotin moiety still bound to the streptavidin.

Figure 8: Mechanism for the cleavage of photocleavable linkers.

Figure 9: The structures of ddNTPs linked to photocleavable (PC) biotin. Any of the four ddNTPs (ddATP, ddCTP, ddGTP, ddTTP) can be used with any of the shown linkers.

Figure 10: The synthesis of ddATP-Linker-II-PC-Biotin. PC = photocleavable.

Figure 11: Schematic for capturing a DNA fragment terminated with a ddNTP on a surface and then for freeing the ddNTP and DNA fragment. The dideoxynucleotide (ddNTP), which is on one end of the

Detailed Description Of The Invention

The following definitions are presented as an aid in understanding this invention.

The standard abbreviations for nucleotide bases are used as follows: adenine (A), cytosine (C), guanine (G), thymine (T), and uracil (U).

This invention is directed to a method for sequencing DNA by detecting the identity of a dideoxynucleotide incorporated to the 3′ end of a DNA sequencing fragment using mass spectrometry, which comprises:

(a) attaching a chemical moiety via a linker to a dideoxynucleotide to produce a labeled dideoxynucleotide;

(b) terminating a DNA sequencing reaction with the labeled dideoxynucleotide to generate a labeled DNA sequencing fragment, wherein the DNA sequencing fragment has a 3′ end and the chemical moiety is attached via the linker to the 3′ end of the DNA sequencing fragment;

(c) capturing the labeled DNA sequencing fragment on a surface coated with a compound that specifically interacts with the chemical moiety attached via the linker to the DNA sequencing fragment, thereby capturing the DNA sequencing fragment;

(d) washing the surface to remove any non-bound component;

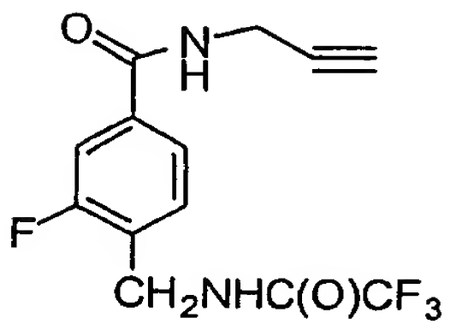
(e) freeing the DNA sequencing fragment from the surface; and

In one embodiment, the chemical moiety is attached via a different linker to different dideoxynucleotides. In one embodiment, the different linkers increase mass separation between different labeled DNA sequencing fragments and thereby increase mass spectrometry resolution.

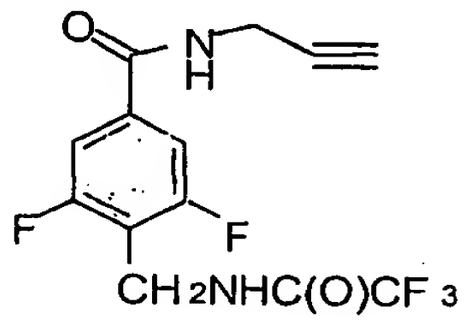
In one embodiment, the dideoxynucleotide is selected from the group consisting of 2',3'-dideoxyadenosine 5'-triphosphate (ddATP), 2',3'-dideoxyguanosine 5'-triphosphate (ddGTP), 2',3'-dideoxycytidine 5'-triphosphate (ddCTP), and 2',3'-dideoxythymidine 5'-triphosphate (ddTTP).

In different embodiments of the methods described herein, the interaction between the chemical moiety attached via the linker to the DNA sequencing fragment and the compound on the surface comprises a biotin-streptavidin interaction, a phenylboronic acid-salicylhydroxamic acid interaction, or an antigen-antibody interaction.

In one embodiment, the step of freeing the DNA sequencing fragment from the surface comprises disrupting the interaction between the chemical moiety attached via the linker to the DNA sequencing fragment and the compound on the surface. In different embodiments, the interaction is disrupted by a means selected from the group consisting of one or more of a physical means, a chemical means, a physical chemical means, heat, and light. In one embodiment, the interaction is disrupted by ultraviolet light. In different embodiments, the



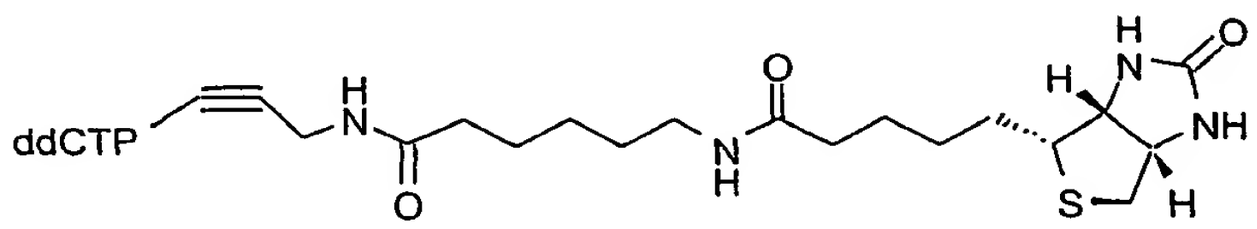
and



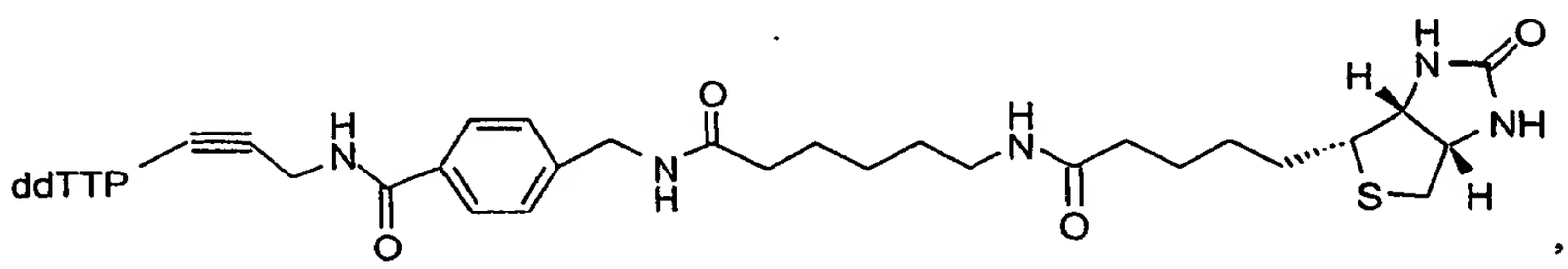
In one embodiment, a plurality of different labeled dideoxynucleotides is used to generate a plurality of different labeled DNA sequencing fragments. In one embodiment, a plurality of different linkers is used to increase mass separation between different labeled DNA sequencing fragments and thereby increase mass spectrometry resolution.

In one embodiment, the chemical moiety comprises biotin, the labeled dideoxynucleotide is a biotinylated dideoxynucleotide, the labeled DNA sequencing fragment is a biotinylated DNA sequencing fragment, and the surface is a streptavidin-coated solid surface. In one embodiment, the biotinylated dideoxynucleotide is selected from the group consisting of ddATP-11-biotin, ddCTP-11-biotin, ddGTP-11-biotin, and ddTTP-16-biotin.

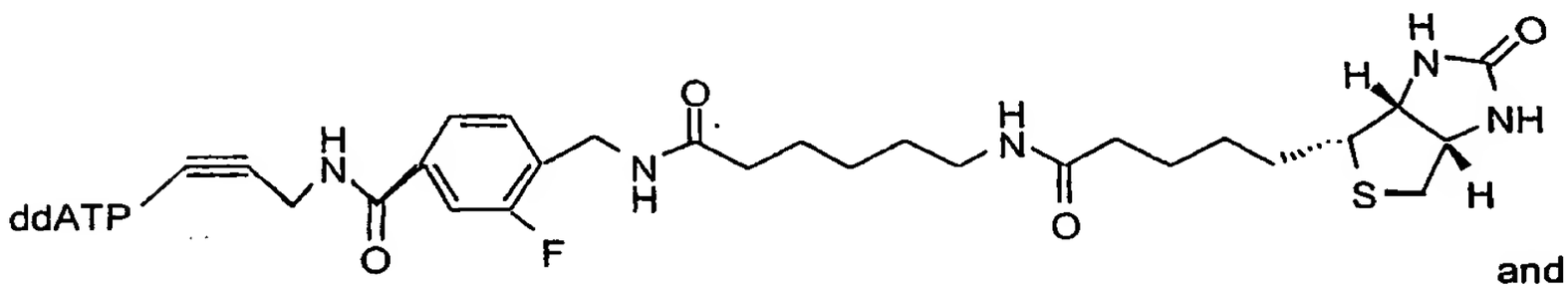
In one embodiment, the biotinylated dideoxynucleotide is selected from the group consisting of:



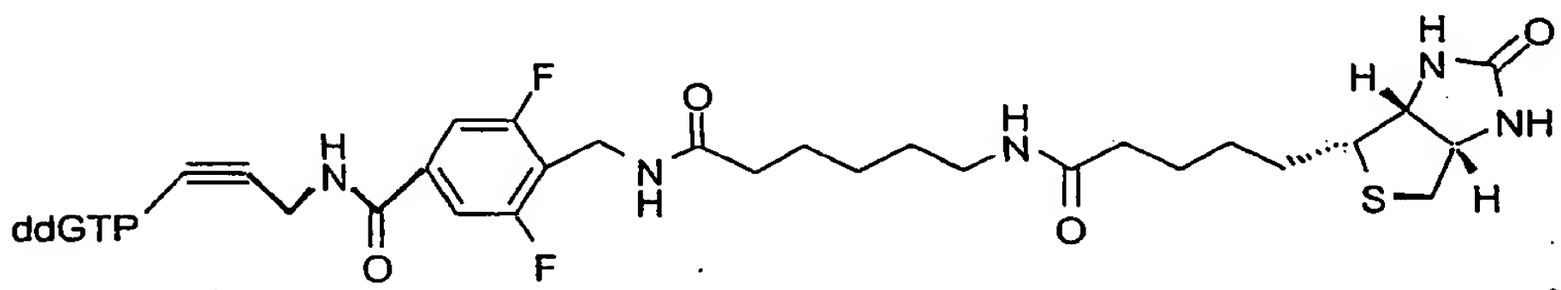
,



,

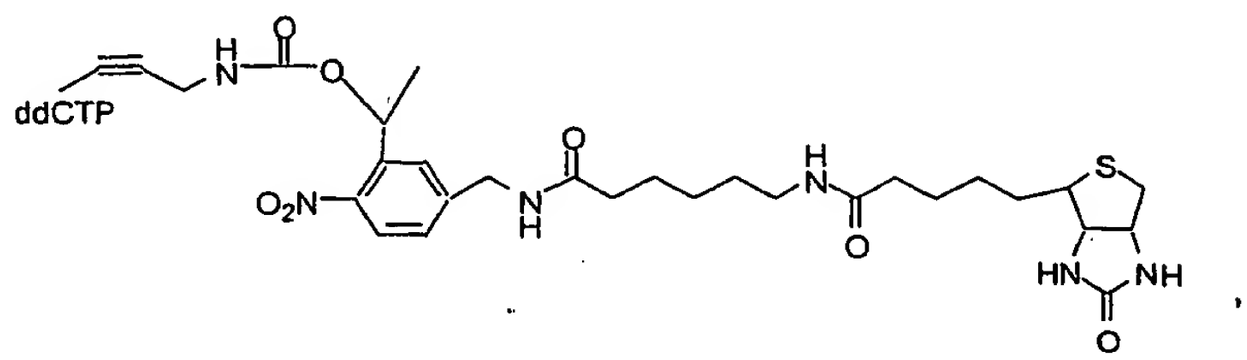


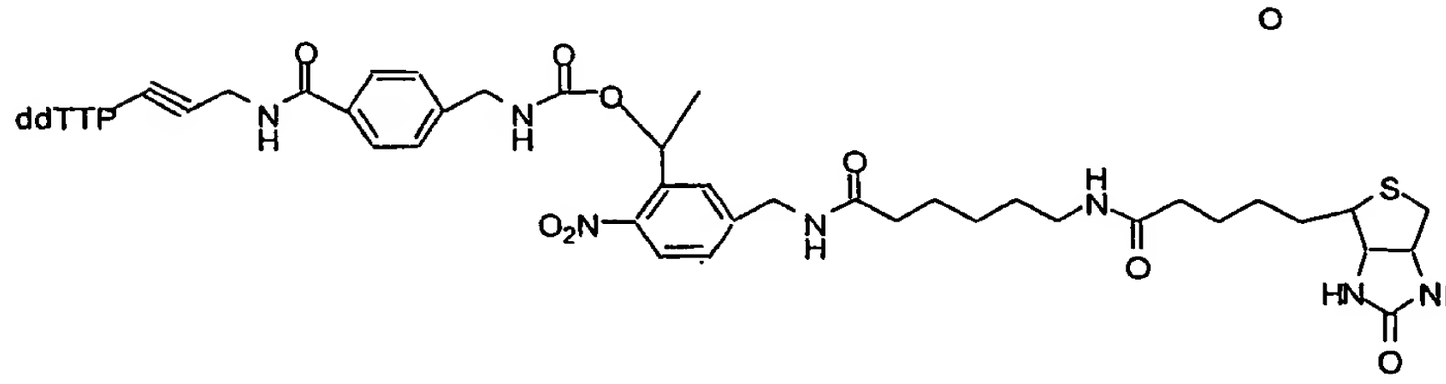
and

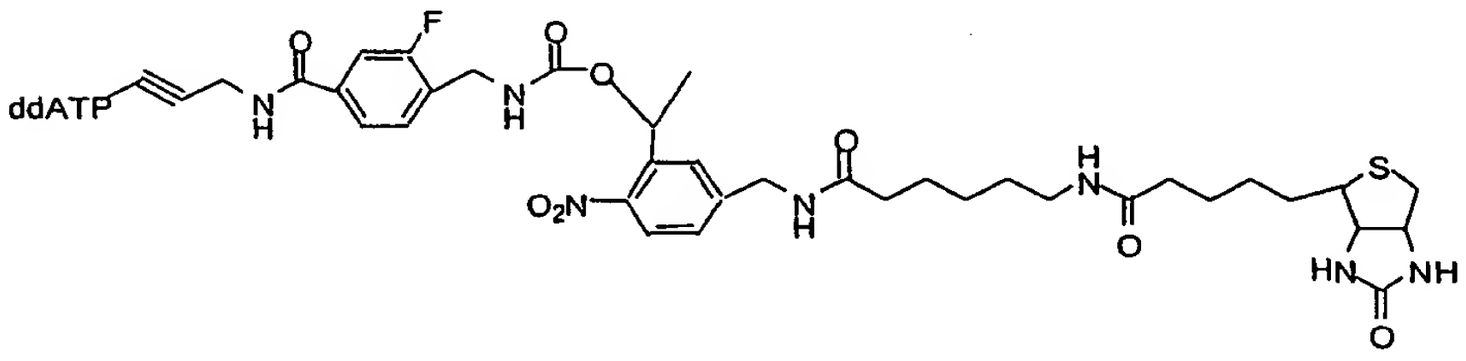


5

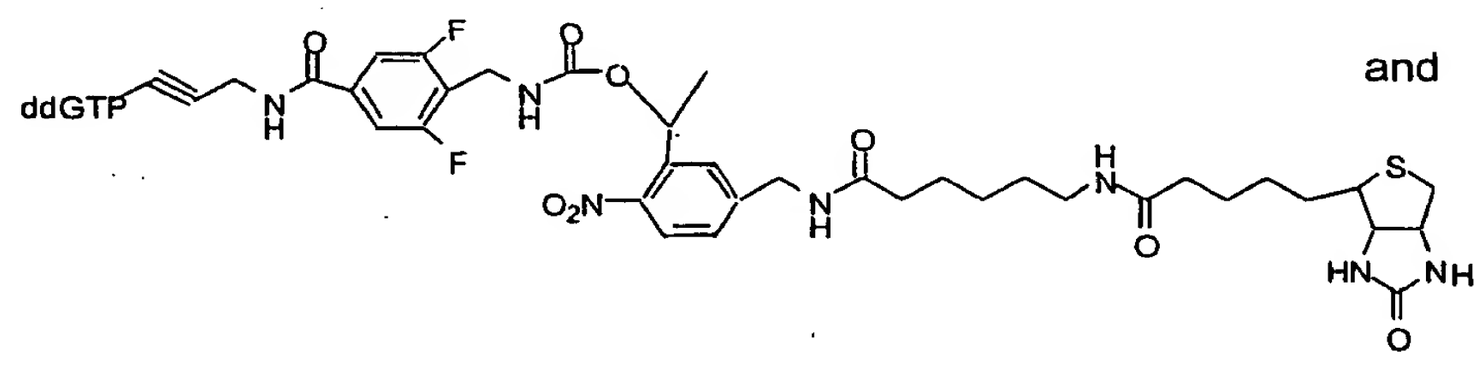
In one embodiment, the biotinylated dideoxynucleotide is selected from the group consisting of:







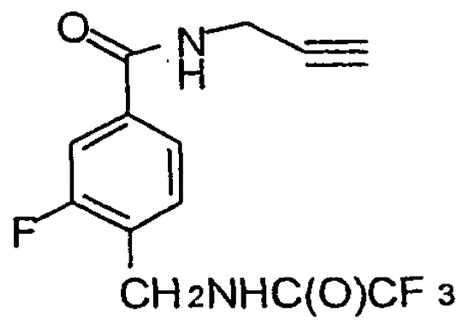
and



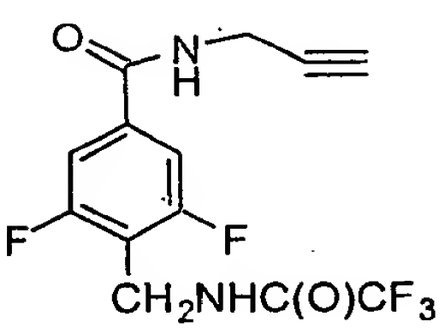
In one embodiment, the streptavidin-coated solid surface is a streptavidin-coated magnetic bead or a streptavidin-coated silica glass.

In one embodiment of the method, steps (b) to (e) are performed in a single container or in a plurality of connected containers.

In one embodiment, the mass spectrometry is matrix-



and



In different embodiments, the linker can comprise a chain structure, or a structure comprising one or more rings, or a structure comprising a chain and one or more rings.

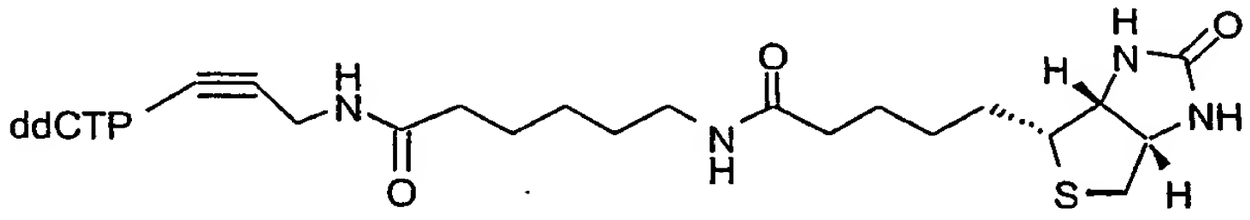
In different embodiments, the linker is cleavable by a means selected from the group consisting of one or more of a physical means, a chemical means, a physical chemical means, heat, and light. In one embodiment, the linker is cleavable by ultraviolet light. In different embodiments, the linker is cleavable by ammonium hydroxide, formamide, or a change in pH (-log H^+ concentration).

In different embodiments of the linker, the chemical moiety comprises biotin, streptavidin, phenylboronic

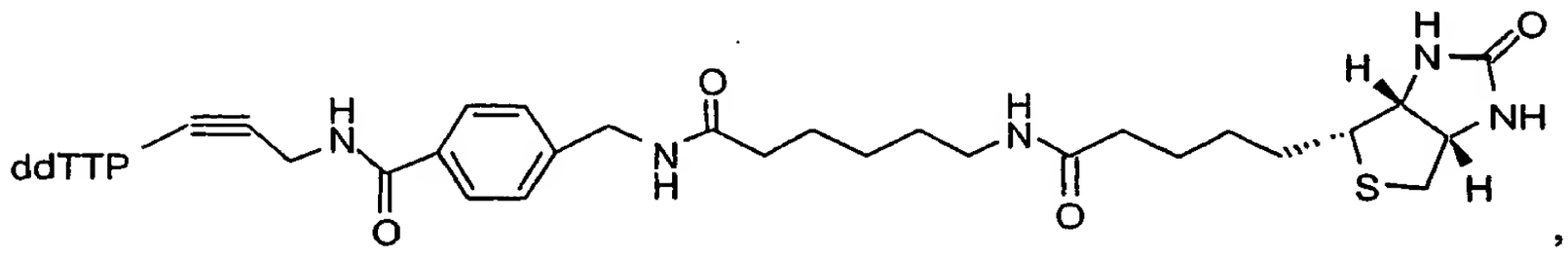
chemical means, a physical chemical means, heat, and light. In one embodiment, the linker is cleavable by ultraviolet light. In different embodiments, the linker is cleavable by ammonium hydroxide, formamide, or a change in pH (-log H^+ concentration).

In different embodiments of the labeled dideoxynucleotide, the chemical moiety comprises biotin, streptavidin, phenylboronic acid, salicylhydroxamic acid, an antibody, or an antigen.

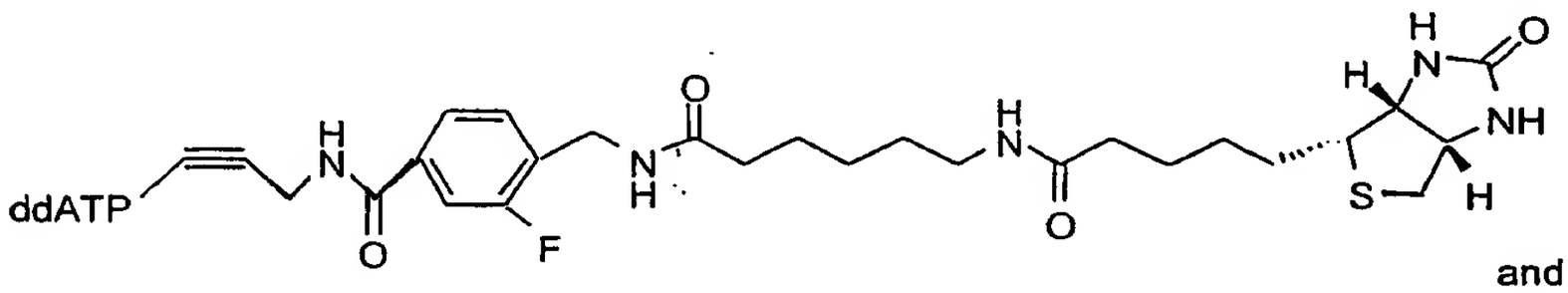
In one embodiment, the labeled dideoxynucleotide is selected from the group consisting of:



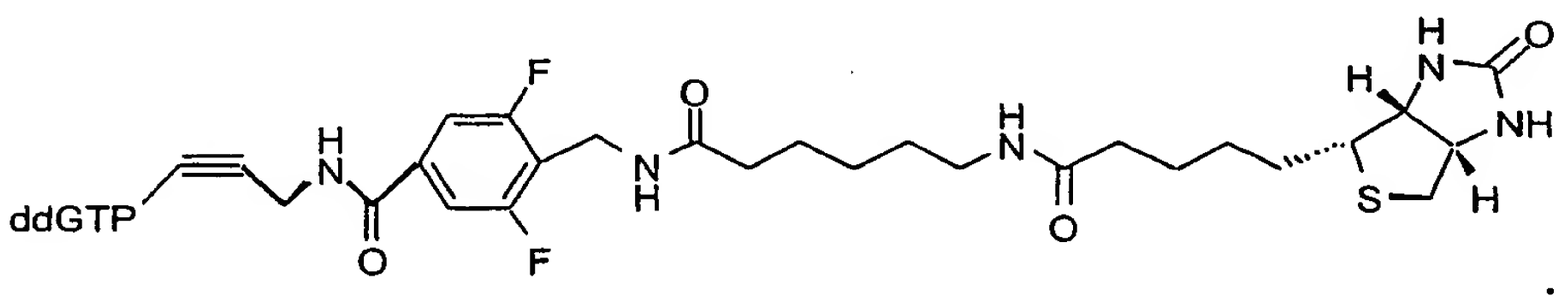
,



,

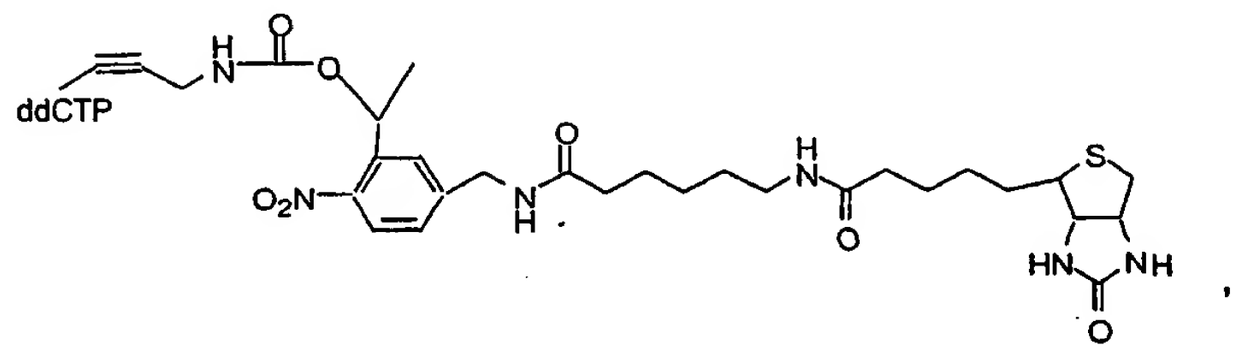


and

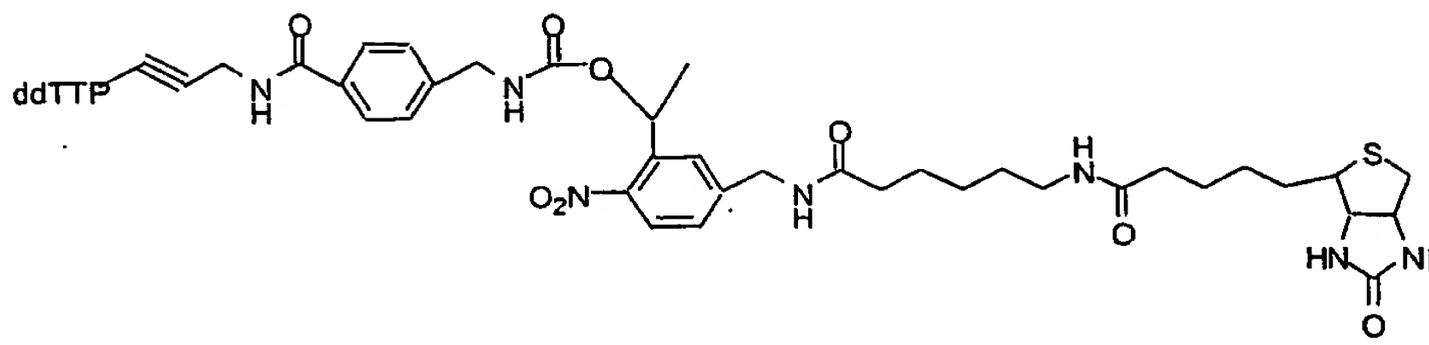


.

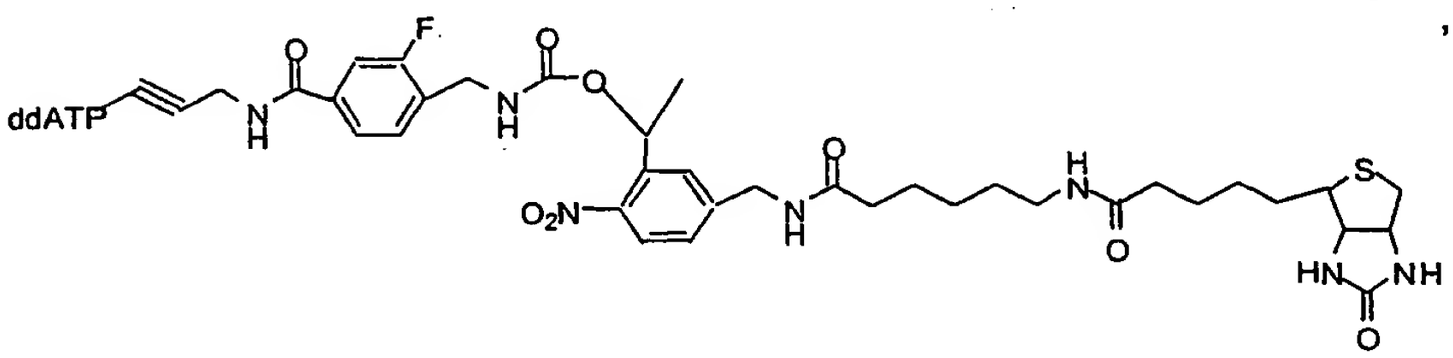
In one embodiment, the labeled dideoxynucleotide is selected from the group consisting of:



,

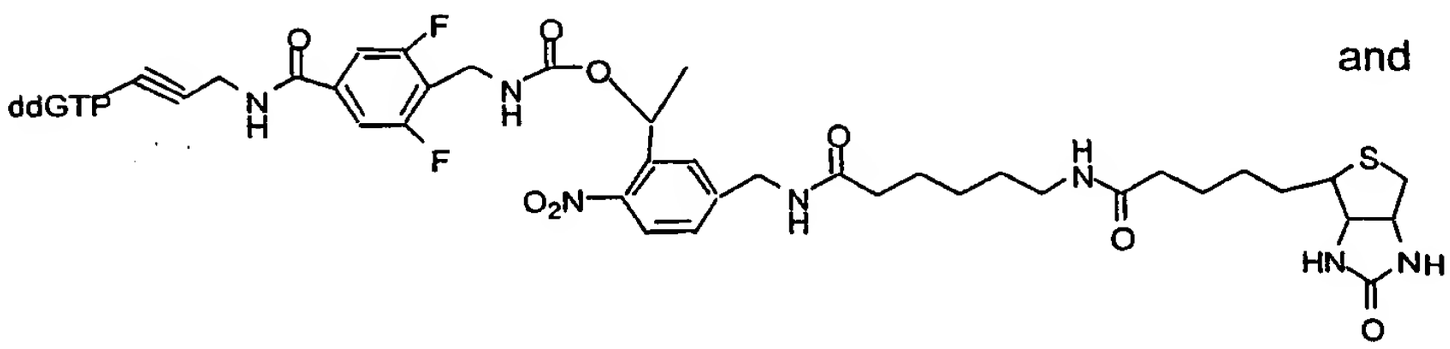


,



,

and



.

The invention provides the use of any of the labeled dideoxynucleotide described herein in DNA sequencing using mass spectrometry, wherein the linker increases mass separation between different labeled dideoxynucleotides and increases mass spectrometry resolution.

In one embodiment, the labeled dideoxynucleotide has

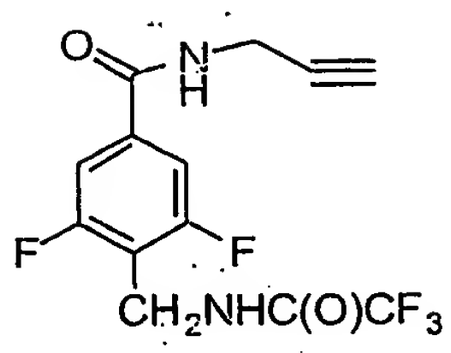
embodiment, the biotinylated moiety is a biotinylated DNA sequencing fragment.

In one embodiment, the chemical moiety can be freed from the surface by disrupting the interaction between the chemical moiety and the compound coating the surface. In different embodiments, the interaction can be disrupted by a means selected from the group consisting of one or more of a physical means, a chemical means, a physical chemical means, heat, and light. In different embodiments, the interaction can be disrupted by ammonium hydroxide, formamide, or a change in pH (-log H^+ concentration).

In one embodiment, the chemical moiety is attached via a linker to another chemical compound. In one embodiment, the other chemical compound is a DNA sequencing fragment. In one embodiment, the linker is cleavable by a means selected from the group consisting of one or more of a physical means, a chemical means, a physical chemical means, heat, and light. In one embodiment, the channel is transparent to ultraviolet light and the linker is cleavable by ultraviolet light. Cleaving the linker frees the DNA sequencing fragment or other chemical compound from the chemical moiety which remains captured on the surface.

The invention provides a multi-channel system which comprises a plurality of any of the single channel systems disclosed herein. In one embodiment, the channels are in a chip. In one embodiment, the multi-channel system comprises 96 channels in a chip.

and



This invention will be better understood from the Experimental Details which follow. However, one skilled in the art will readily appreciate that the specific methods and results discussed are merely illustrative of the invention as described more fully in the claims which follow thereafter.

phenylboronic acid-salicylhydroxamic acid (Bergseid et al. 2000) and antigen-antibody systems.

As illustrated schematically in Figure 1, DNA template, deoxynucleotides (dNTPs) (A, C, G, T) and biotinylated dideoxynucleotides (ddNTP-biotin) (A-b, C-b, G-b, T-b), primer, and DNA polymerase are combined in one tube. After polymerase extension and termination reactions, a series of DNA sequencing fragments with different lengths are generated. The sequencing reaction mixture is then incubated for a few minutes with a streptavidin coated solid phase. Only the DNA sequencing fragments that are terminated with biotinylated dideoxynucleotide at the 3' end are captured on the solid phase. Excess primers, false terminated DNA fragments (fragments terminated at dNTPs instead of ddNTPs), enzymes and all other components from the sequencing reaction are washed away. The biotinylated DNA sequencing fragments are then cleaved off the solid phase by disrupting the interaction between biotin and streptavidin to obtain a pure set of DNA sequencing fragments. The interaction between biotin and streptavidin can be disrupted using, for example, ammonium hydroxide, formamide, or a change in pH. The DNA sequencing fragments are then mixed with matrix (3-hydroxy-picolinic acid) and loaded into a mass spectrometer to produce accurate mass spectra of the DNA sequencing fragments. Since each type of nucleotide has a unique molecular mass, the mass difference between adjacent peaks on the mass spectra gives the sequence identity of the nucleotides.

template (SEQ ID NO: 1) and 13 bp primer (SEQ ID NO: 2):

5'-ACTTTTTACTGTTCGATCCCTGCATCTCAGAGCTCGCTATTCCGAGCTTACACGT-3'
Template
3'-TAAGGCTCGAATG-5'
Primer

Four commercially available biotinylated dideoxynucleotides ddATP-11-biotin, ddGTP-11-biotin, ddCTP-11-biotin and ddTTP-11-biotin (New England Nuclear, Boston) were used to produce the sequencing ladder that was generated all in one tube using the cycle sequencing procedure. It can be seen from Figure 2 that very clean sequence peaks are obtained on the mass spectra, with the first peak being primer extended by one biotinylated dideoxynucleotide. Furthermore, excess primer in the sequencing reaction is completely removed and no false stopped peaks are detected. The base identity of A and G can be identified unambiguously in Figure 2. Since the mass difference between the commercially available ddCTP-11-Biotin and ddTTP-11-biotin is one dalton and the resolution is only within about 3 daltons in the mass detector for DNA fragments, C and T cannot be differentiated in Figure 2. The data shows that by capturing/releasing DNA sequencing fragments with the biotin located on the 3' dideoxy terminators, clean sequencing ladders that are free from any other contaminants can be obtained. Further improvement of the procedure requires the use of biotinylated ddTTPs that have large mass differences in comparison to ddCTP-11-biotin. To achieve this, ddTTP-16-biotin is

Sample preparation is performed in one tube by executing the sequencing reactions with biotinylated ddNTPs, regular dNTPs, DNA polymerase, and reaction buffer. The sample is then placed in a thermocycler for 30 cycles to create extension fragments. Streptavidin beads are then added to the sample and incubated to allow the biotin-streptavidin complex to form. The beads are collected by placing the reaction tube in a magnet and thoroughly washing them with an ammonium acetate solution to remove all impurities such as false stops, primers, and salts. Dilute ammonium hydroxide solution is then used to dissociate the biotin streptavidin complex at 60 °C (Jurinke et. al., 1997). Once this complex is dissociated, the solution is placed back in the magnet to separate the beads out of solution. The supernatant is collected, added to a matrix solution of 3-hydroxy-picolinic acid (Aldrich), and allowed to crystallize for analysis by a Perkin Elmer Voyager DE MALDI-TOF mass spectrometer. The resulting spectrum is assigned according to the positions of the various peaks.

on the bases in the nucleotides, even with bulky energy transfer fluorescent dyes, still allows efficient incorporation of the modified nucleotides into the DNA strand by DNA polymerase (Rosenblum et al. 1997, Zhu et al. 1994). Thus, the ddNTPs-Linker-11-biotin can be incorporated into the growing strand by the polymerase in DNA sequencing reactions.

Larger mass separations will greatly aid in longer read lengths where signal intensity is smaller and resolution is lower. The smallest mass difference between two individual bases is over three times as great in the mass tagged biotinylated ddNTPs compared to normal ddNTPs and more than double that achieved by the standard biotinylated ddNTPs as shown in Table 1. Three 4-aminomethyl benzoic acid derivatives **Linker I**, **Linker II** and **Linker III** are designed as mass tags as well as linkers for bridging biotin to the corresponding dideoxynucleotides. The synthesis of **Linker II** (Figure 5) is described here to illustrate the synthetic procedure. 3-Fluoro-4-aminomethyl benzoic acid that can be easily prepared via published procedures (Maudling et al. 1983; Rolla 1982) is first protected with trifluoroacetic anhydride, then converted to N-hydroxysuccinimide (NHS) ester with disuccinimidylcarbonate in the presence of diisopropylethylamine. The resulting NHS ester is subsequently coupled with commercially available propargylamine to form the desired compound, **Linker II**. Using an analogous procedure, **Linker I** and **Linker III** can be easily constructed.

released into solution by cleaving the photocleavable linker with ultraviolet (UV) light, while the biotin remains attached to the streptavidin that is covalently bound to the surface. The pure DNA fragments can then be crystallized in matrix solution and analyzed by mass spectrometry. It is advantageous to cleave the biotin moiety since it contains sulfur which has several relatively abundant isotopes. The rest of the DNA fragments and linkers contain only carbon, nitrogen, hydrogen, oxygen, fluorine and phosphorous, whose dominant isotopes are found with a relative abundance of 99% to 100%. This allows high resolution mass spectra to be obtained. The photocleavage mechanism (Olejnik et al. 1995, 1999) is shown in Figure 8. Upon irradiation with ultraviolet light at 300-350 nm, the light sensitive o-nitroaromatic carbonamide functionality on DNA fragment 1 is cleaved, producing DNA fragment 2, PC-biotin and carbon dioxide. The partial chemical linker remaining on DNA fragment 2 is stable for detection by mass spectrometry.

Four new biotinylated ddNTPs disclosed here, ddCTP-PC-Biotin, ddTTP-Linker I-PC-Biotin, ddATP-Linker II-PC-Biotin and ddGTP-Linker III-PC-Biotin are shown in Figure 9. These compcunds are synthesized by a similar chemistry as shown for the synthesis of ddATP-Linker II-11-Biotin in Figure 6. The only difference is that in the final coupling step NHS-PC-LC-Biotin (Pierce, Rockford IL) is used, as shown in Figure 10. The photocleavable linkers disclosed here allow the use of solid phase capturable terminators

disrupted by a means selected from the group consisting of one or more of a physical means, a chemical means, a physical chemical means, heat, and light. In one embodiment, ultraviolet light can be used to cleave the cleavable linker. Chemical means include, but are not limited to, ammonium hydroxide (Jurinke et. al., 1997), formamide, or a change in pH (-log H^+ concentration) of the solution.

V. High density streptavidin-coated, porous silica channel system.

Streptavidin coated magnetic beads are not ideal for using the photocleavable biotin capture and release process for DNA sequencing fragments, since they are not transparent to UV light. Therefore, the photocleavage reaction is not efficient. For efficient capture of the biotinylated sequencing fragments, a high-density surface coated with streptavidin is essential. It is known that the commercially available 96-well streptavidin coated plates cannot provide a sufficient surface area for efficient capture of the biotinylated DNA fragments. Disclosed in this application is a new porous silica channel system designed to overcome this limitation.

To increase the surface area available for solid phase capture, porous channels are coated with a high density of streptavidin. Ninety-six (96) porous silica glass channels can be etched into a silica chip (Figure 12). The surfaces of the channels are modified to contain streptavidin as shown in Figure 13. The channel is first treated with 0.5 M NaOH,

The fragment solution is then driven out of the channel and into a collection plate. After matrix solution is added, the samples are spotted on a chip and allowed to crystallize for detection by MALDI-TOF mass spectrometry. The purification cassette is cleaned by chemically cleaving the biotin-streptavidin linkage, and is then washed and reused.

VI. Validation of the Mass Spectrometry DNA Sequencing System Using Synthetic DNA Templates and PCR Templates Generated from Genomic DNA.

To validate the sequencing technology disclosed here, a synthetic DNA template can be synthesized which mimics a portion of the human immunodeficiency virus type 1 protease gene. The sequence of the template (SEQ ID NO: 3) and that of the sequencing primer (SEQ ID NO: 4) are shown below (Schmit et al. 1996):

5'-TAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATGGTCCAGGTCGTG-3'
Template
3'-CCAGGTCCAGCAC-5'
Primer

The tumor suppressor gene p53 can also be used as a model system. The p53 gene is one of the most frequently mutated genes in human cancer (O'Connor et al. 1997). Since most of the p53 mutation hot spots are clustered within exons 5-8, this region of the p53 gene is selected as a sequencing target. A synthetic sequencing template containing a portion of the sequences from exon 7 and exon 8 of the p53 gene and an appropriate primer can be prepared:

Template: 5'-CATGTGTAACAGTTCCTGCATGGGCGGCATGAACCCGAGG

References

Bergseid M, Baytan AR, Wiley JP, Ankener WM, Stolowitz, Hughs KA, Chestnut JD (Nov. 2000) Small-molecule base chemical affinity system for the purification of proteins. *BioTechniques* **29**: 1126-1133.

Bowling JM, Bruner KL, Cmarik JL, Tibbetts C. (1991) Neighboring nucleotide interactions during DNA sequencing gel electrophoresis. *Nucleic Acids Res.* **19**: 3089-3097.

Burgess K, Cook D. (2000) *Chemical Reviews.* **100**: 2047-2060.

Chee M, Yang R, Hubbell E, Berno, A, Huang, XC., Stern D, Winkler, J, Lockhart DJ, Morris M S, Fodor, SP. (1996) Accessing genetic information with high-density DNA arrays. *Science* **274**: 610-614.

Chiu NH, Tang K, Yip P, Braun A, Koster H, Cantor CR. (2000) Mass spectrometry of single-stranded restriction fragments captured by an undigested complementary sequence. *Nucleic Acids Res.* **28**: E31.

Fei Z, Ono T, Smith LM. (1998) MALDI-TOF mass spectrometric typing of single nucleotide polymorphisms with mass-tagged ddNTPs. *Nucleic Acids Res.* **26**: 2827-2828.

Fu DJ, Tang K, Braun A, Reuter D, Darnhofer-Demar B, Little DP, O'Donnell MJ, Cantor CR, Koster H. (1998)

ammonium hydroxide and applications in MALDI-TOF mass spectrometry. *Anal. Chem.* 69: 904-910.

Kheterpal I, Scherer J, Clark SM, Radhakrishnan A, Ju J, Ginther CL, Sensabaugh GF, Mathies RA. (1996) DNA Sequencing Using a Four-Color Confocal Fluorescence Capillary Array Scanner. *Electrophoresis* 17: 1852-1859.

Langer PR, Waldrop AA, Ward DC. (1981) Enzymatic synthesis of biotin-labeled polynucleotides: novel nucleic acid affinity probes. *Proc. Natl. Acad. Sci. USA*. 78: 6633-6637.

Lee LG, Connell CR, Woo SL, Cheng RD, Mcardle BF, Fuller CW, Halloran ND, Wilson RK. (1992) DNA sequencing with dye-labeled terminators and T7 DNA - polymerase-effect of dyes and dNTPs on incorporation of dye-terminators and probability analysis of termination fragments. *Nucleic Acids Res.* 20: 2471-2483.

Maudling DR, Lotts KD, Robinson SA. (1983) New procedure for making 2-(chloromethyl)-4-nitrotoluene. *J. Org. Chem.* 48: 2938.

Monforte JA, Becker CH (1997) High-throughput DNA analysis by time-of-flight mass spectrometry. *Nat Med.* 3(3): 360-362.

O'Connor PM, Jackman J, Bae I, Myers TG, Fan S, Mutoh M, Scudiero DA, Monks A, Sausville EA, Weinstein JN, Friend S, Fornace AJ Jr, Kohn KW. (1997) Characterization of the p53 tumor suppressor

Rolla F. (1982) Sodium-borohydride reactions under phase-transfer conditions. - reduction of azides to amines. *J. Org. Chem.* **47**: 4327-4329.

Rosenblum BB, Lee LG, Spurgeon SL, Khan SH, Menchen SM, Heiner CR, Chen SM. (1997) New dye-labeled terminators for improved DNA sequencing patterns. *Nucleic Acids Res.* **25**: 4500-4504.

Roses A. (2000) Pharmacogenetics and the practice of medicine. *Nature.* **405**: 857-865.

Roskey MT, Juhasz P, Smirnov IP, Takach EJ, Martin SA, Haff LA. (1996) DNA sequencing by delayed extraction-matrix-assisted laser desorption/ ionization time of flight mass spectrometry. *Proc. Natl. Acad. Sci. USA.* **93**: 4724-4729.

Salas-Solano O, Carrilho E, Kotler L, Miller AW, Goetzinger W, Sosic Z, Karger BL, (1998) Routine DNA sequencing of 1000 bases in less than one hour by capillary electrophoresis with replaceable linear polyacrylamide solutions. *Anal. Chem.* **70**: 3996-4003.

Sanger F, Nickeln S and Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**: 5463-5467

Schmit J.C, Ruiz L, Clotet B, Raventos A, Tor J, Leonard J, Desmyter J. De Clercq E, Vandamme AM. (1996) Resistance-related mutations in the HIV-1 protease gene of patients treated for 1 year with the

Tong, X., Smith LM (1993) Solid Phase Purification in Automated DNA Sequencing. *DNA Sequence-J.DNA Sequencing and Mapping* **4**: 151-162.

Velculescu VE, Zhang, I, Vogelstein, B. and Kinzler KW (1995) Serial Analysis of Gene Expression. *Science* **270**: 484-487.

Woolley AT, Mathies RA. (1994) Ultra-high-speed DNA fragment separations using microfabricated capillary array electrophoresis chips. *Proc. Natl. Acad. Sci. USA*. **91**: 11348-11352.

Wu KJ, Steding A, Becker CH. (1993) Matrix-assisted laser desorption time-of-flight mass spectrometry of oligonucleotides using 3-hydroxypicolinic acid as an ultraviolet-sensitive matrix. *Rapid Commun Mass Spectrom*. **7**: 142-146.

Xu L, Bian N, Wang Z, Abdel-Baky S, Pillai S, Magiera D, Murugaiah V, Giese RW, Wang P, O'Keeffe T, Abushamaa H, Kutney L, Church G, Carson S, Smith D, Park M, Wronka J, Laukien F. (1997) Electrophore mass tag dideoxy DNA sequencing. *Anal. Chem.* **69**: 3595-3602.

Yamakawa H, Ohara O. (1997) A DNA cycle sequencing reaction that minimizes compressions on automated fluorescent sequencers. *Nucleic. Acids. Res.* **25**: 1311-1312.

Zhu Z, Chao J, Yu H, Waggoner AS. (1994) Directly labeled DNA probes using fluorescent nucleotides with

What is claimed is:

1. A method for sequencing DNA by detecting the identity of a dideoxynucleotide incorporated to the 3′ end of a DNA sequencing fragment using mass spectrometry, which comprises:

(a) attaching a chemical moiety via a linker to a dideoxynucleotide to produce a labeled dideoxynucleotide;

(b) terminating a DNA sequencing reaction with the labeled dideoxynucleotide to generate a labeled DNA sequencing fragment, wherein the DNA sequencing fragment has a 3′ end and the chemical moiety is attached via the linker to the 3′ end of the DNA sequencing fragment;

(c) capturing the labeled DNA sequencing fragment on a surface coated with a compound that specifically interacts with the chemical moiety attached via the linker to the DNA sequencing fragment, thereby capturing the DNA sequencing fragment;

(d) washing the surface to remove any non-bound component;

(e) freeing the DNA sequencing fragment from the surface; and

(f) analyzing the DNA sequencing fragment using mass spectrometry so as to sequence the DNA.

2. A method for sequencing DNA by detecting the identity of a plurality of dideoxynucleotides incorporated to the 3′ end of different DNA

and the compound on the surface comprises a biotin-streptavidin interaction, a phenylboronic acid-salicylhydroxamic acid interaction, or an antigen-antibody interaction.

5. The method of claim 1 or 2, wherein the step of freeing the DNA sequencing fragment from the surface comprises disrupting the interaction between the chemical moiety attached via the linker to the DNA sequencing fragment and the compound on the surface.

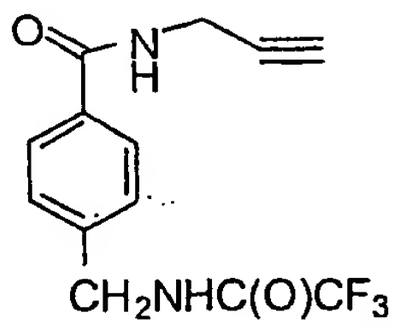
6. The method of claim 5, wherein the interaction is disrupted by a means selected from the group consisting of one or more of a physical means, a chemical means, a physical chemical means, heat, and light.

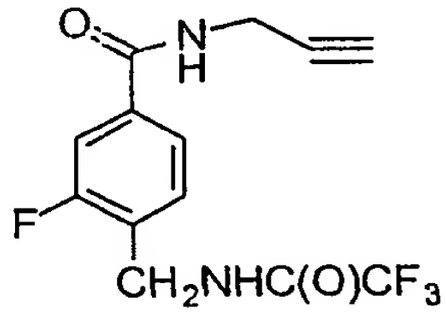
7. The method of claim 1 or 2, wherein the dideoxynucleotide comprises a cytosine or a thymine with a 5-position, or an adenine or a guanine with a 7-position, and the linker is attached to the 5-position of cytosine or thymine or to the 7-position of adenine or guanine.

8. The method of claim 1 or 2, wherein the step of freeing the DNA sequencing fragment from the surface comprises cleaving the linker.

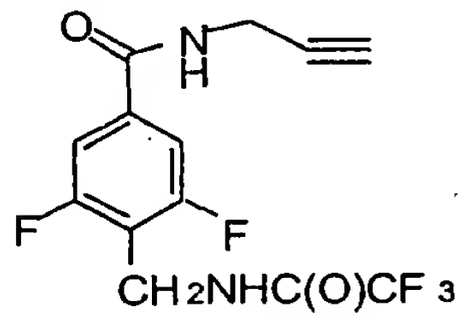
9. The method of claim 8, where the linker is cleaved by a means selected from the group consisting of one or more of a physical means, a

13. The method of claim 12, wherein the linker is selected from the group consisting of:





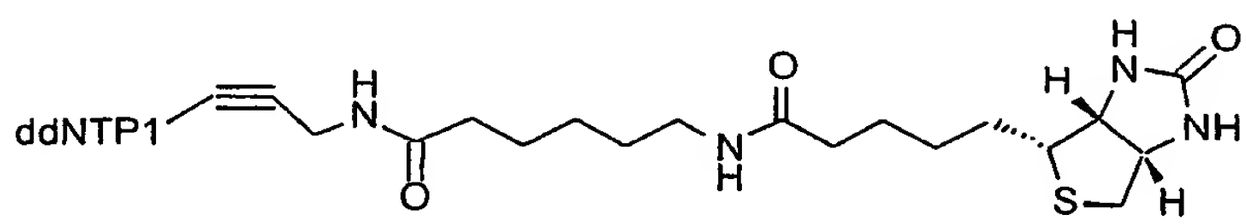
and



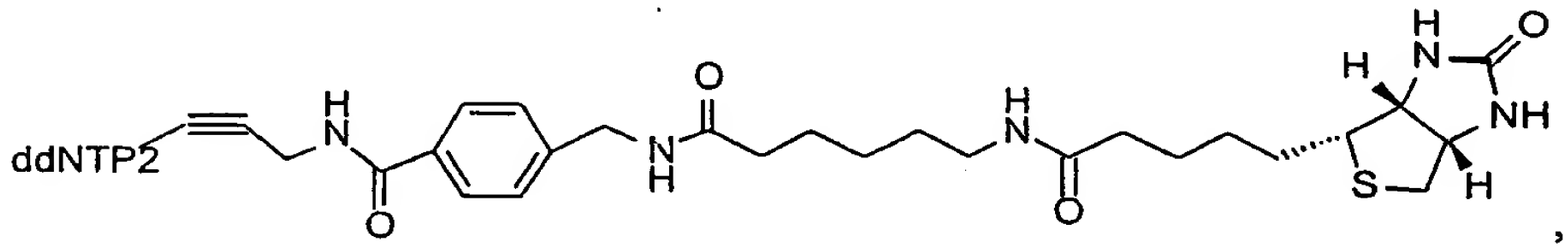
14. The method of claim 1, wherein a plurality of different labeled dideoxynucleotides is used to generate a plurality of different labeled DNA sequencing fragments.

15. The method of claim 3 or 14, wherein a plurality of different linkers is used to increase mass separation between different labeled DNA sequencing fragments and thereby increase mass spectrometry resolution.

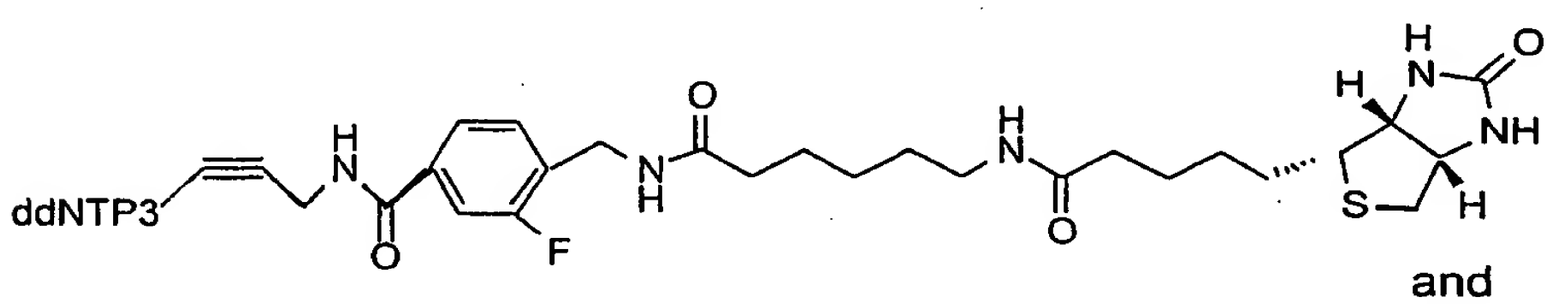
18. The method of claim 16, wherein the biotinylated dideoxynucleotide is selected from the group consisting of:



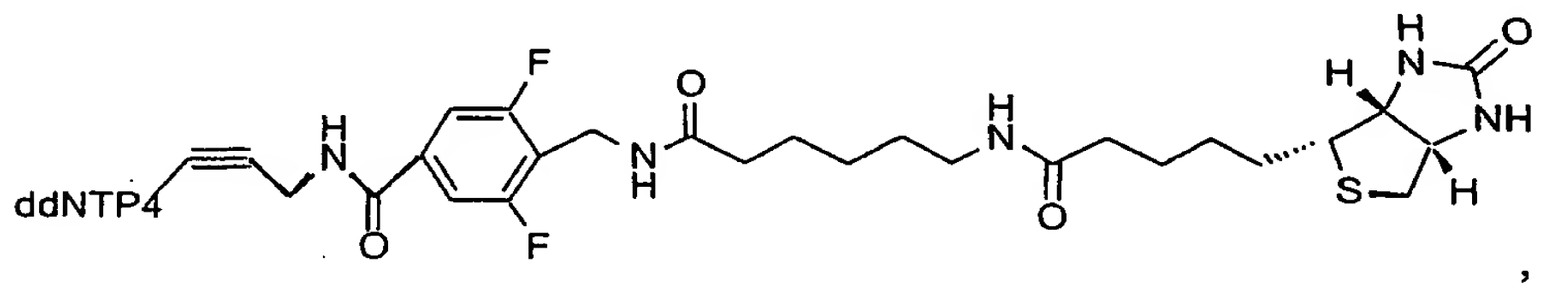
,



,



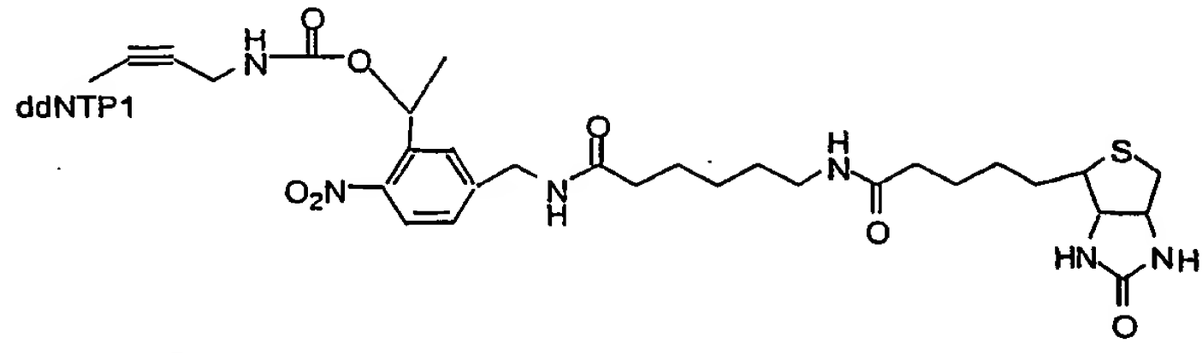
and



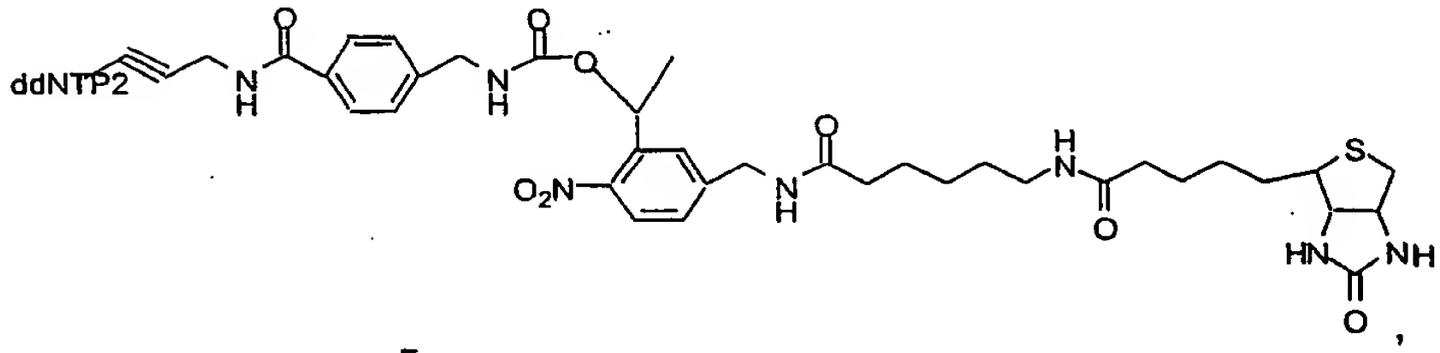
,

wherein ddNTP1, ddNTP2, ddNTP3, and ddNTP4 represent four different dideoxynucleotides.

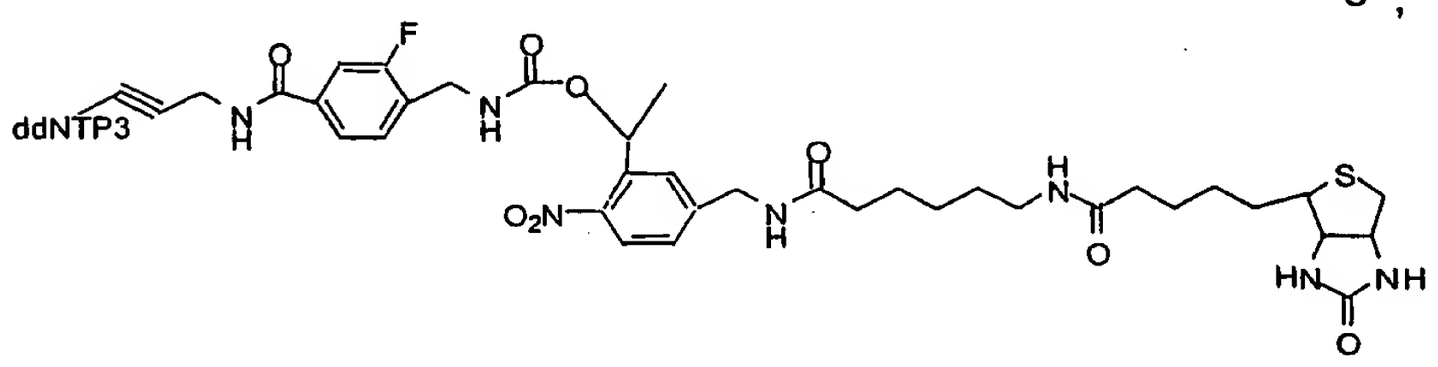
20. The method of claim 16, wherein the biotinylated dideoxynucleotide is selected from the group consisting of:



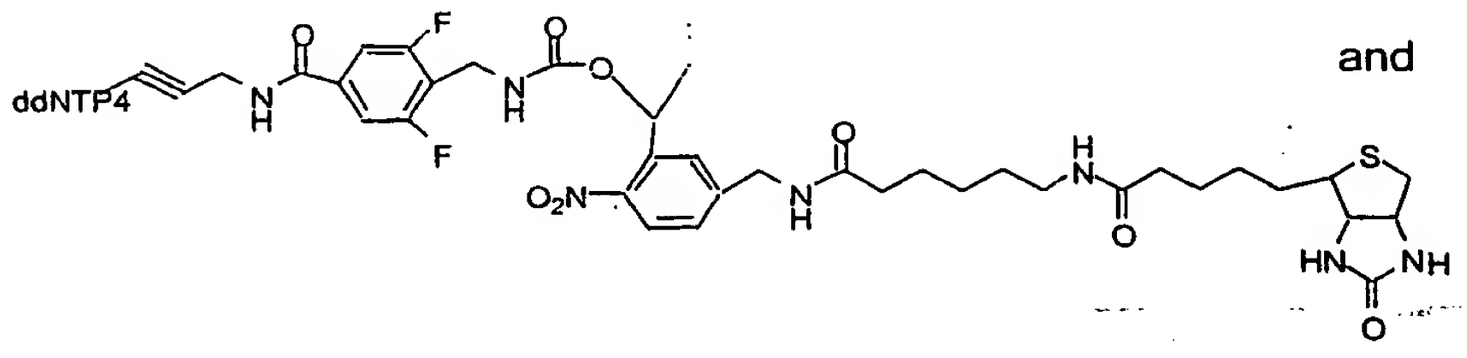
,



,



and



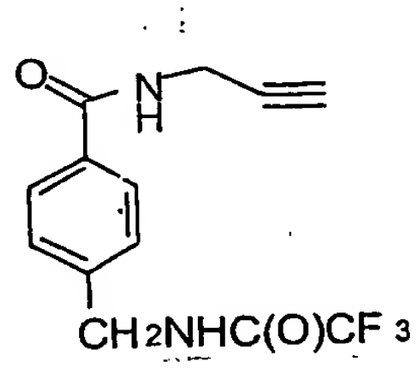
wherein ddNTP1, ddNTP2, ddNTP3, and ddNTP4 represent four different dideoxynucleotides.

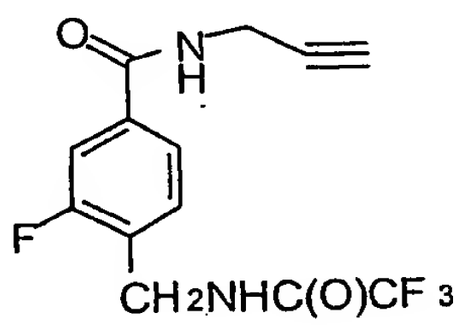
24. Use of the method of claim 1 or 2 for detection of single nucleotide polymorphisms, genetic mutation analysis, serial analysis of gene expression, gene expression analysis, identification in forensics, genetic disease association studies, genomic sequencing, translational analysis, or transcriptional analysis.

25. A linker for attaching a chemical moiety to a dideoxynucleotide, wherein the linker comprises a derivative of 4-aminomethyl benzoic acid.

26. The linker of claim 25, wherein the linker comprises one or more fluorine atoms.

27. The linker of claim 26, wherein the linker is selected from the group consisting of:





and

dideoxynucleotides and increases mass spectrometry resolution.

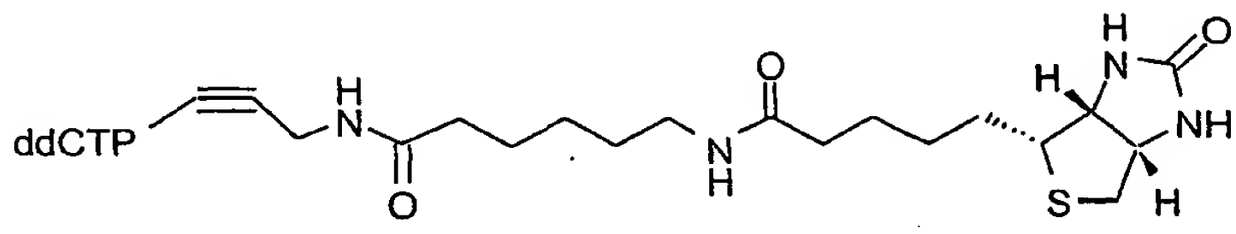
33. A labeled dideoxynucleotide, which comprises a chemical moiety attached via a linker to a 5-position of cytosine or thymine or to a 7-position of adenine or guanine.

34. The labeled dideoxynucleotide of claim 33, wherein the linker is cleavable by a means selected from the group consisting of one or more of a physical means, a chemical means, a physical chemical means, heat, and light.

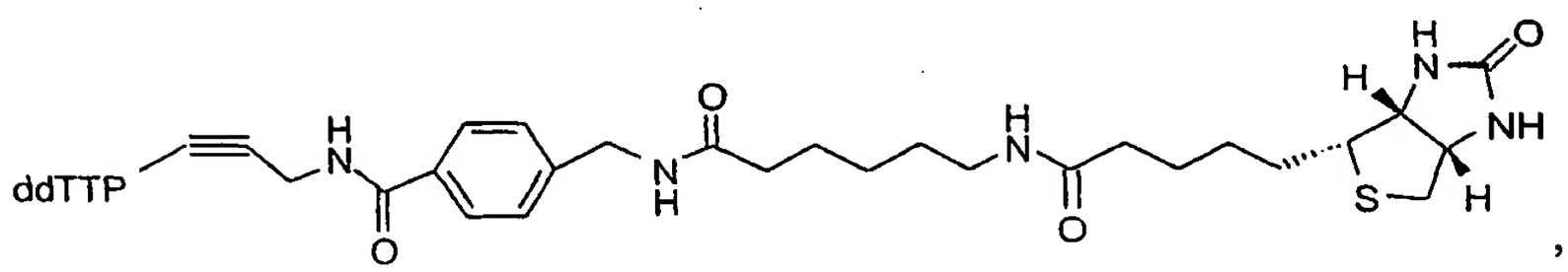
35. The labeled dideoxynucleotide of claim 34, wherein the linker is cleavable by ultraviolet light.

36. The labeled dideoxynucleotide of claim 33, wherein the chemical moiety comprises biotin, streptavidin, phenylboronic acid, salicylhydroxamic acid, an antibody, or an antigen.

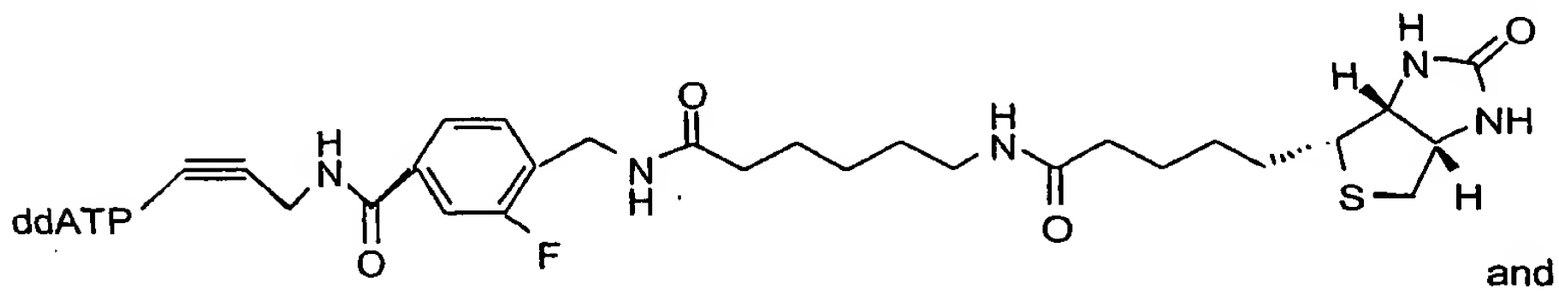
38. The labeled dideoxynucleotide of claim 37, wherein the labeled dideoxynucleotide is selected from the group consisting of:



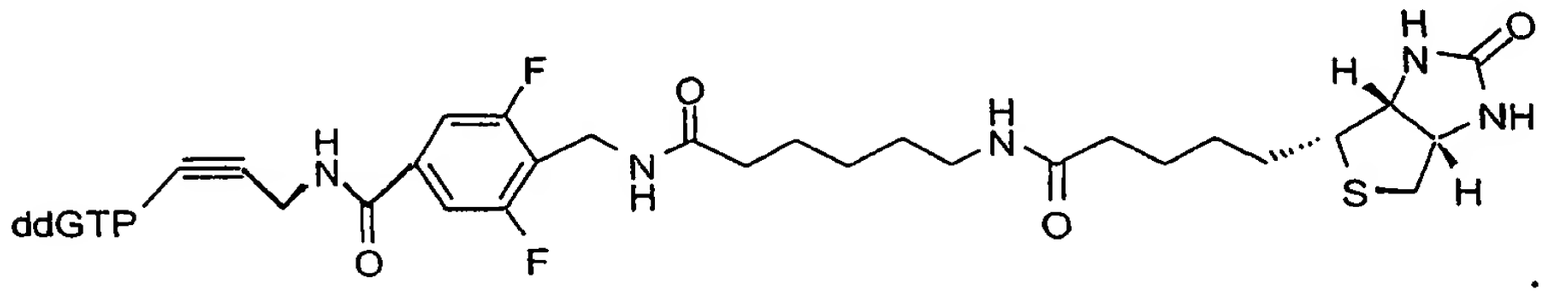
,



,

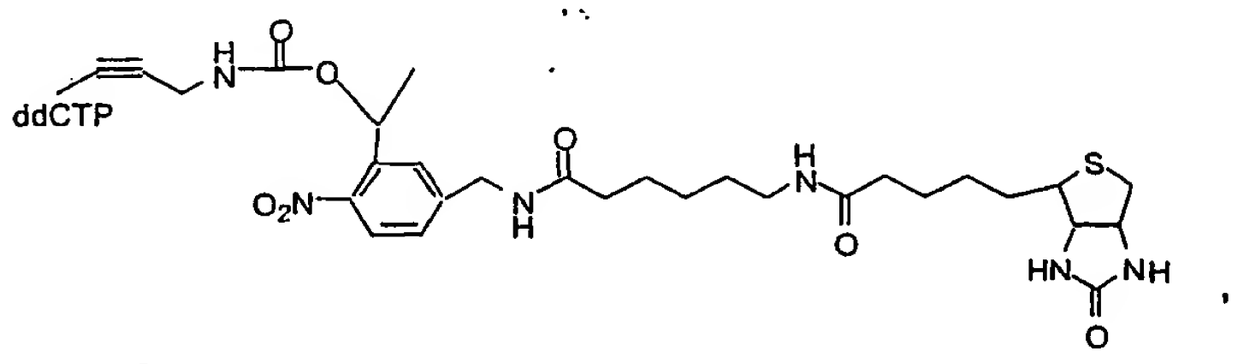


and

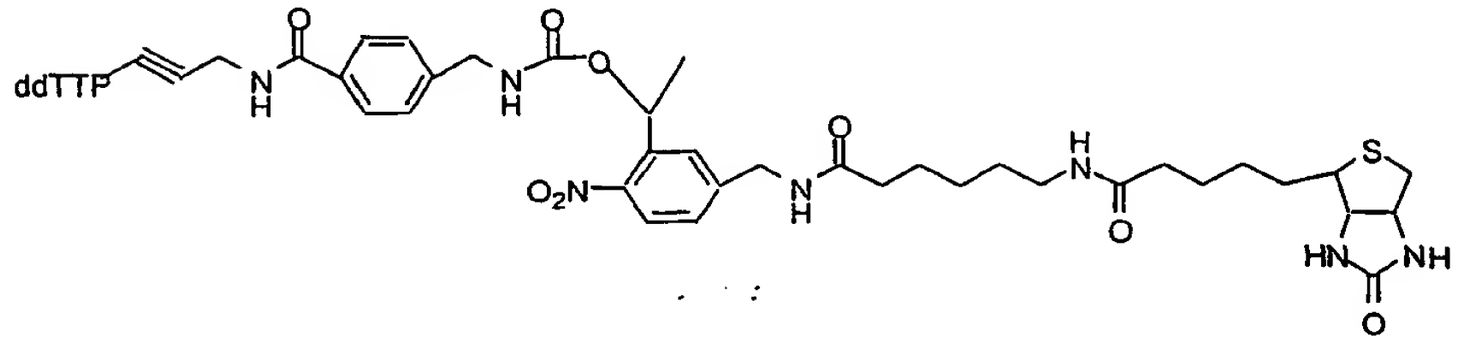


.

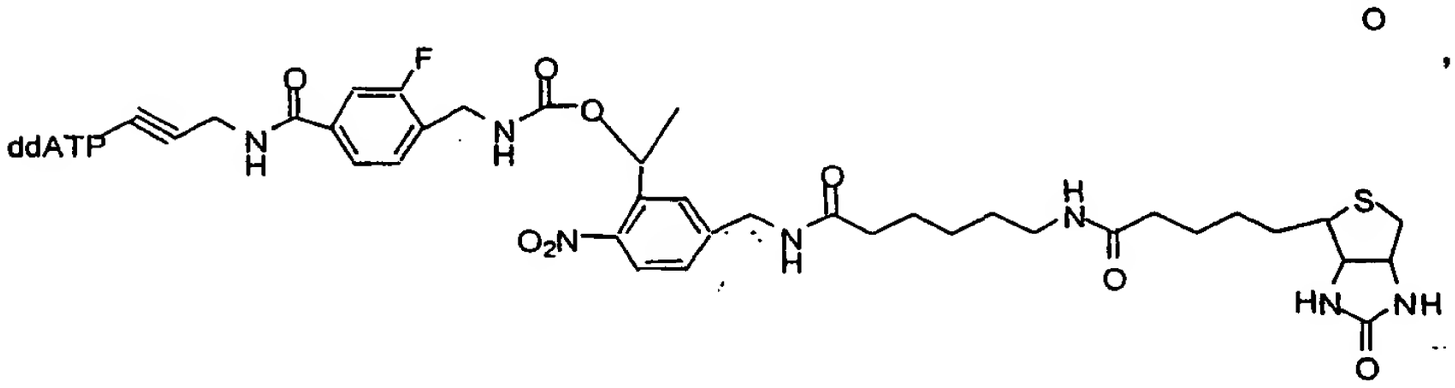
40. The labeled dideoxynucleotide of claim 39, wherein the labeled dideoxynucleotide is selected from the group consisting of:



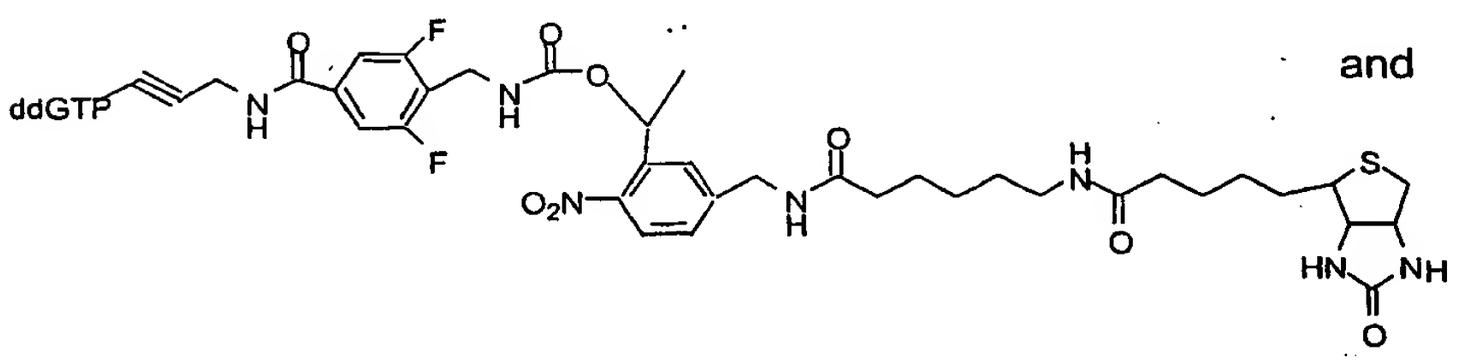
,



,



,



and

41. Use of the labeled dideoxynucleotide of claim 33 in DNA sequencing using mass spectrometry, wherein the linker increases mass separation between different labeled dideoxynucleotides and increases mass spectrometry resolution.

47. The system of claim 46, where the interaction can be disrupted by a means selected from the group consisting of one or more of a physical means, a chemical means, a physical chemical means, heat, and light.

48. The system of claim 42, wherein the chemical moiety is attached via a linker to another chemical compound.

49. The system of claim 48, wherein the other chemical compound is a DNA sequencing fragment.

50. The system of claim 48, where the linker is cleavable by a means selected from the group consisting of one or more of a physical means, a chemical means, a physical chemical means, heat, and light.

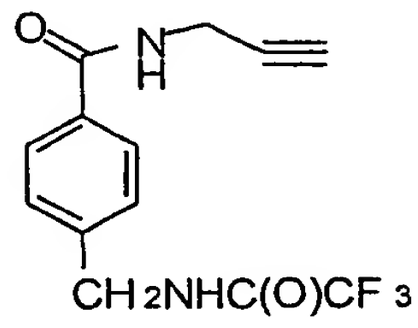
51. The system of claim 50, wherein the channel is transparent to ultraviolet light and the linker is cleavable by ultraviolet light.

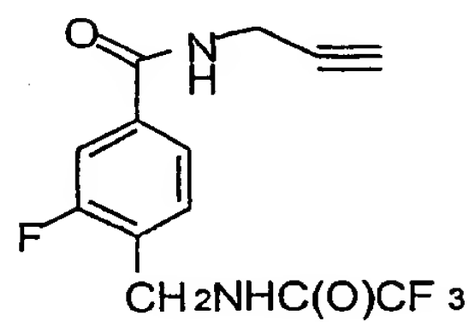
52. A multi-channel system, which comprises a plurality of the system of claim 42.

53. The multi-channel system of claim 52, wherein the channels are in a chip.

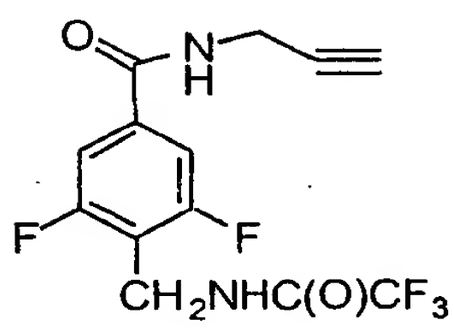
54. The multi-channel system of claim 53, which comprises 96 channels in a chip.

58. The method of claim 57, wherein one or more of the different linkers is selected from the group consisting of:





and



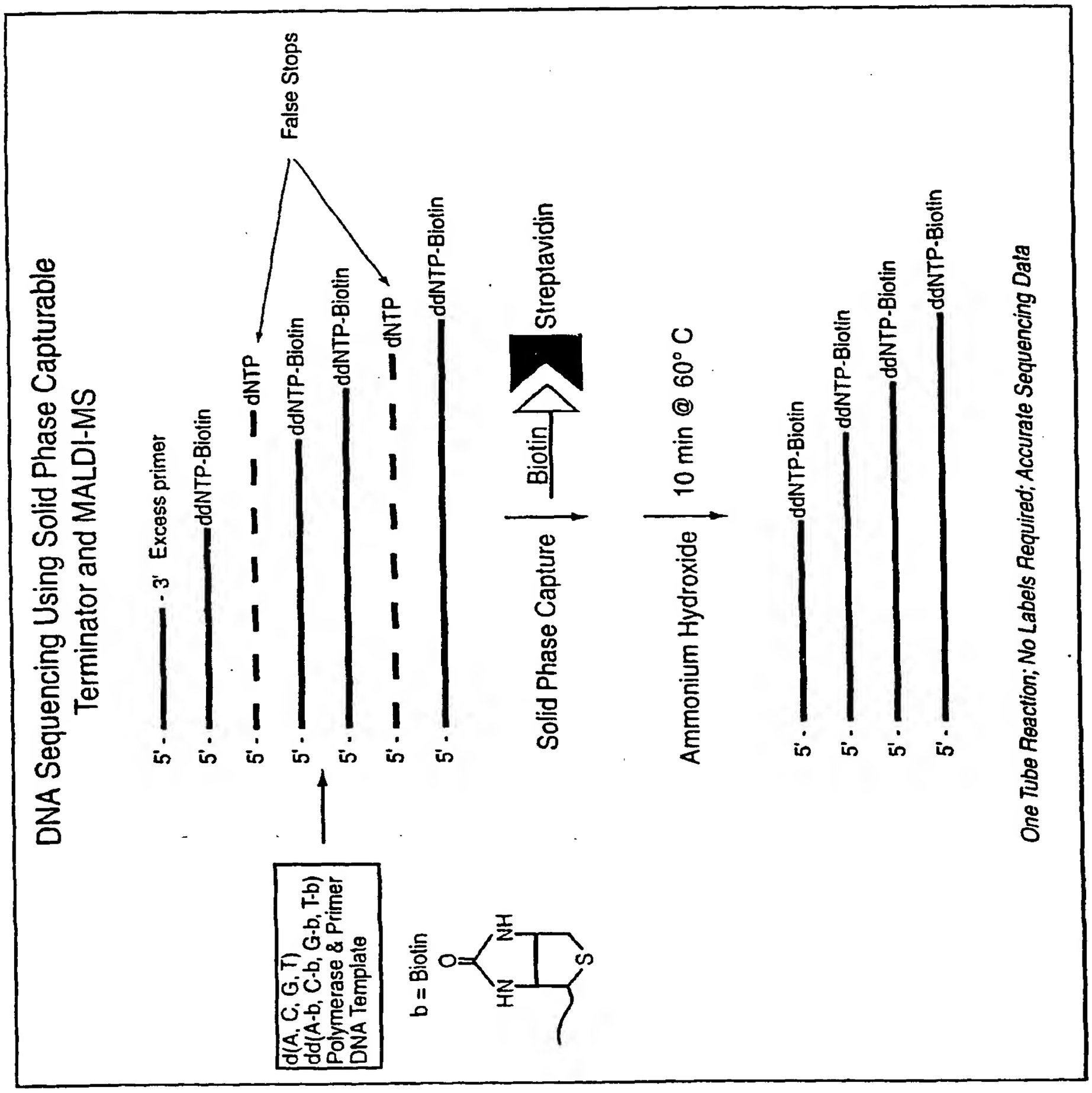

DNA Sequencing Using Solid Phase Capturable Terminator and MALDI-MS
5' - 3' Excess primer
5' - ddNTP-Biotin
5' - dNTP
5' - ddNTP-Biotin
5' - ddNTP-Biotin
5' - dNTP
5' - ddNTP-Biotin
d(A, C, G, T)
dd(A-b, C-b, G-b, T-b)
Polymerase & Primer
DNA Template
False Stops
Solid Phase Capture
Biotin
Streptavidin
Ammonium Hydroxide
10 min @ 60° C
5' - ddNTP-Biotin
5' - ddNTP-Biotin
5' - ddNTP-Biotin
5' - ddNTP-Biotin
b = Biotin
NH
O
HN
S
One Tube Reaction; No Labels Required; Accurate Sequencing Data

FIGURE 1

FIGURE 2

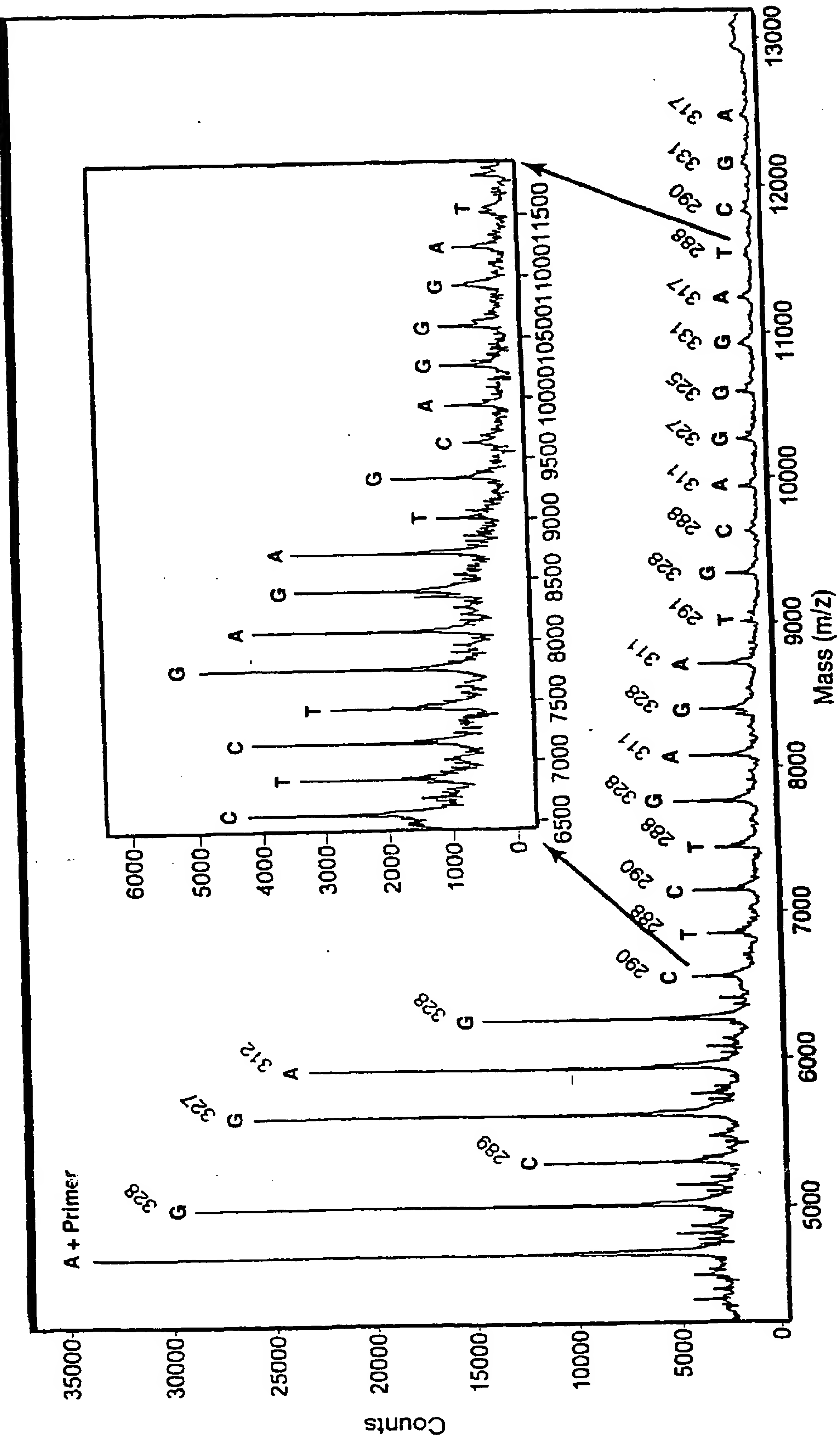

Counts
Mass (m/z)
A + Primer
G 328
C 289
G 327
A 312
G 328
C 290
T 288
C 290
T 288
G 328
A 311
G 328
A 311
T 291
G 328
C 288
A 311
G 327
G 325
G 331
A 317
T 288
C 290
G 331
A 317
C T C T G A G A T G C A G G G A T

FIGURE 3

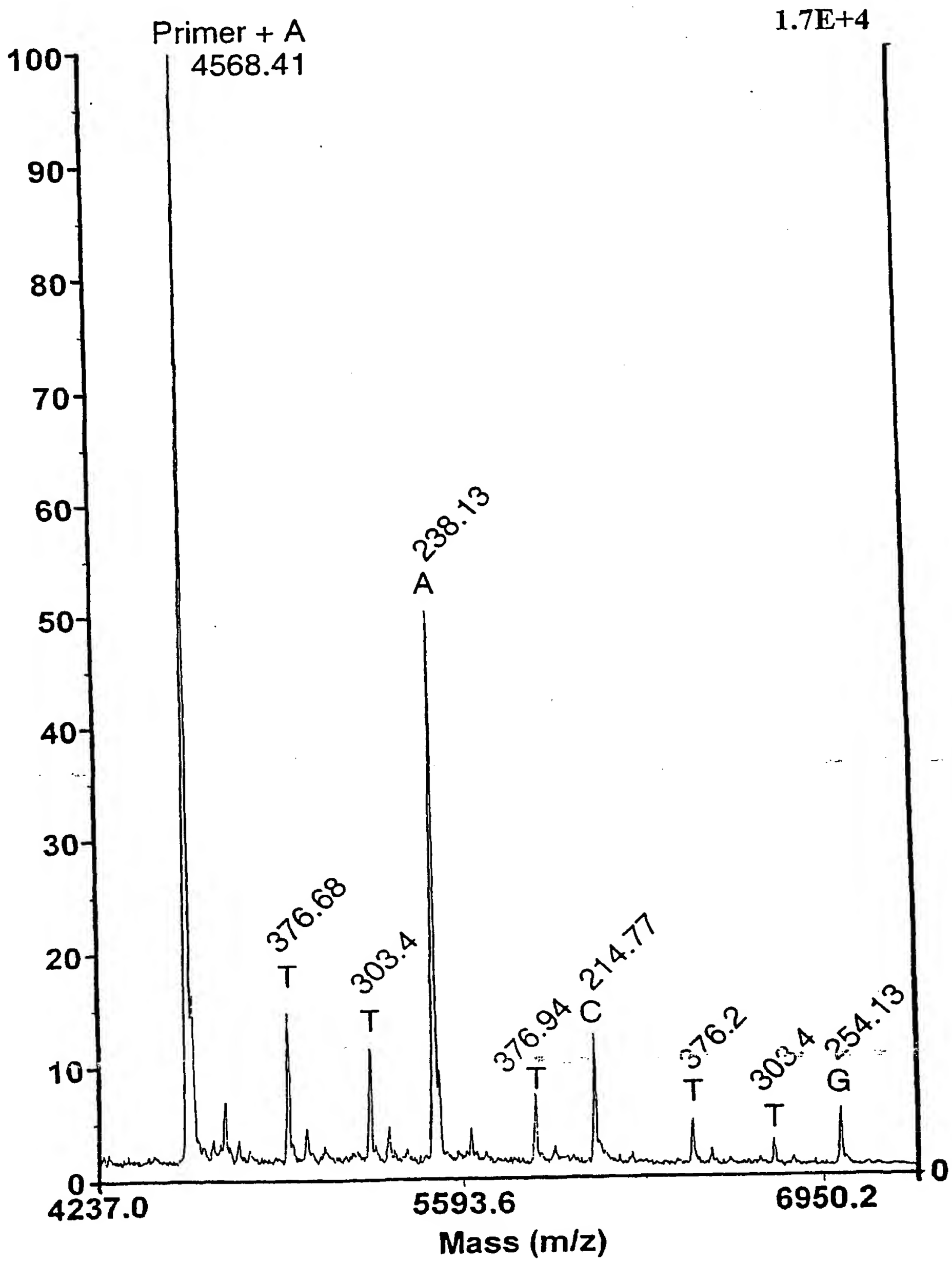
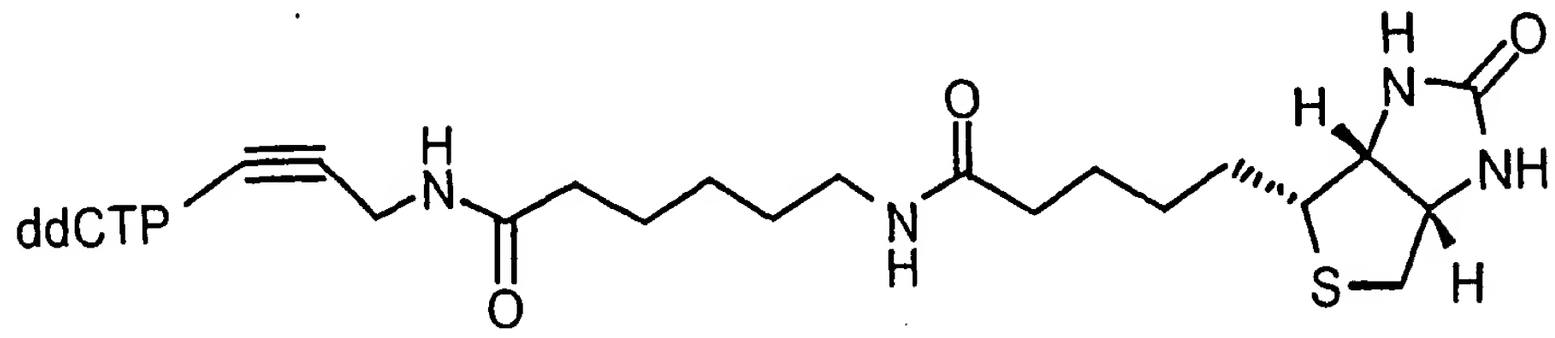
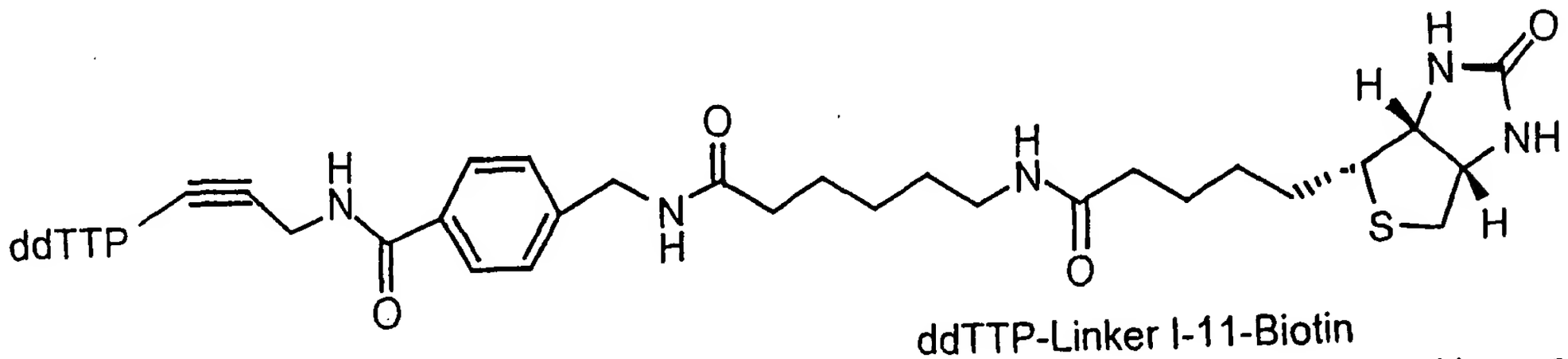

Primer + A
4568.41
1.7E+4
100
90
80
70
60
50
40
30
20
10
0
% Intensity
A 238.13
376.68
T
303.4
T
376.94
T
C 214.77
376.2
T
303.4
T
254.13
G
0
4237.0
5593.6
6950.2
Mass (m/z)

FIGURE 4



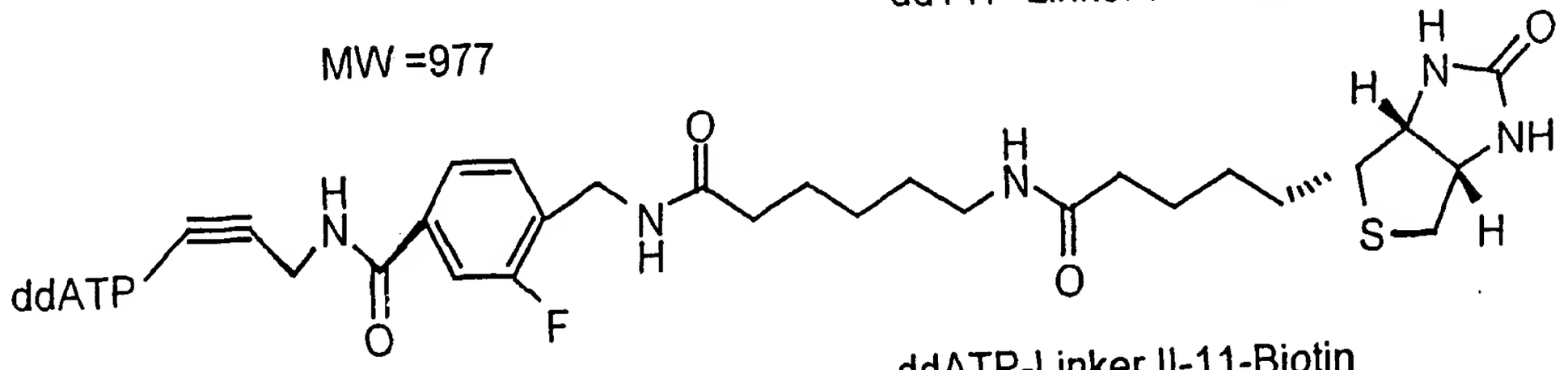
MW = 844

ddCTP-11-Biotin



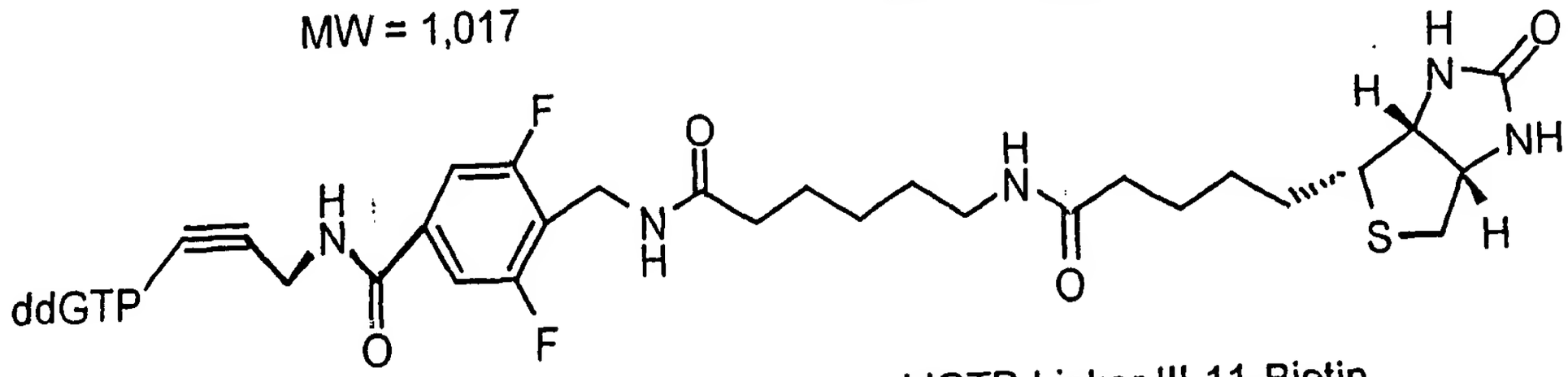
ddTTP-Linker I-11-Biotin

MW =977



ddATP-Linker II-11-Biotin

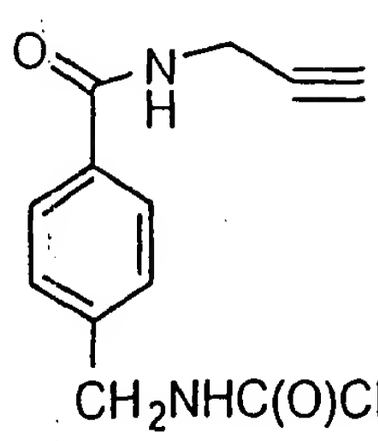
MW = 1,017



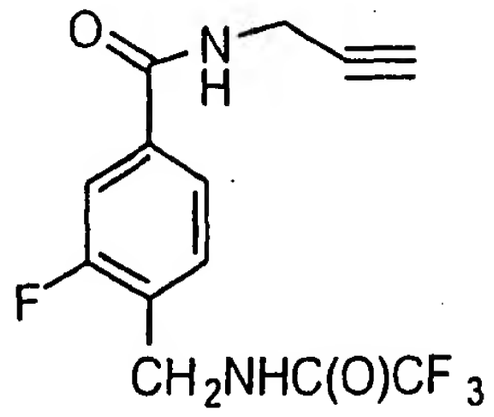
ddGTP-Linker III-11-Biotin

MW = 1,051

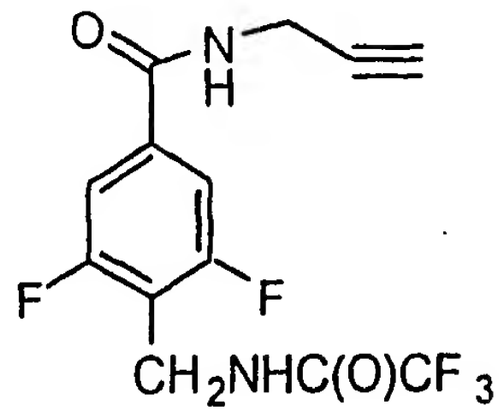
FIGURE 5



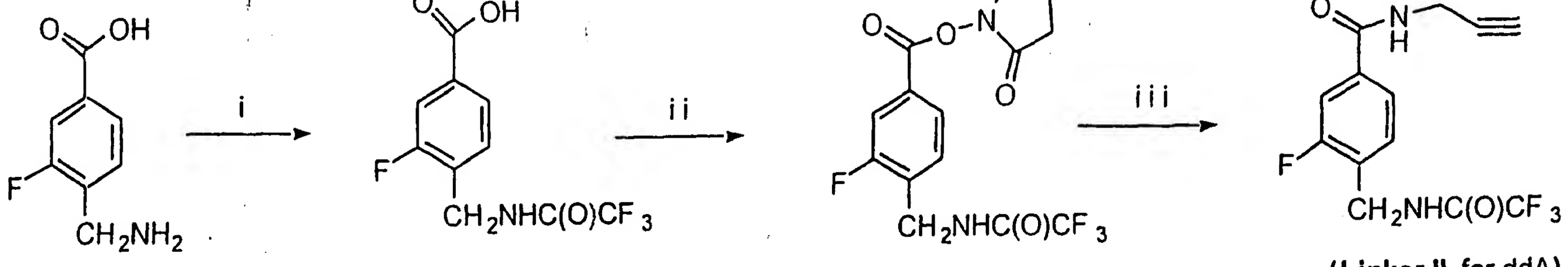
(Linker I, for ddT)



(Linker II, for ddA)



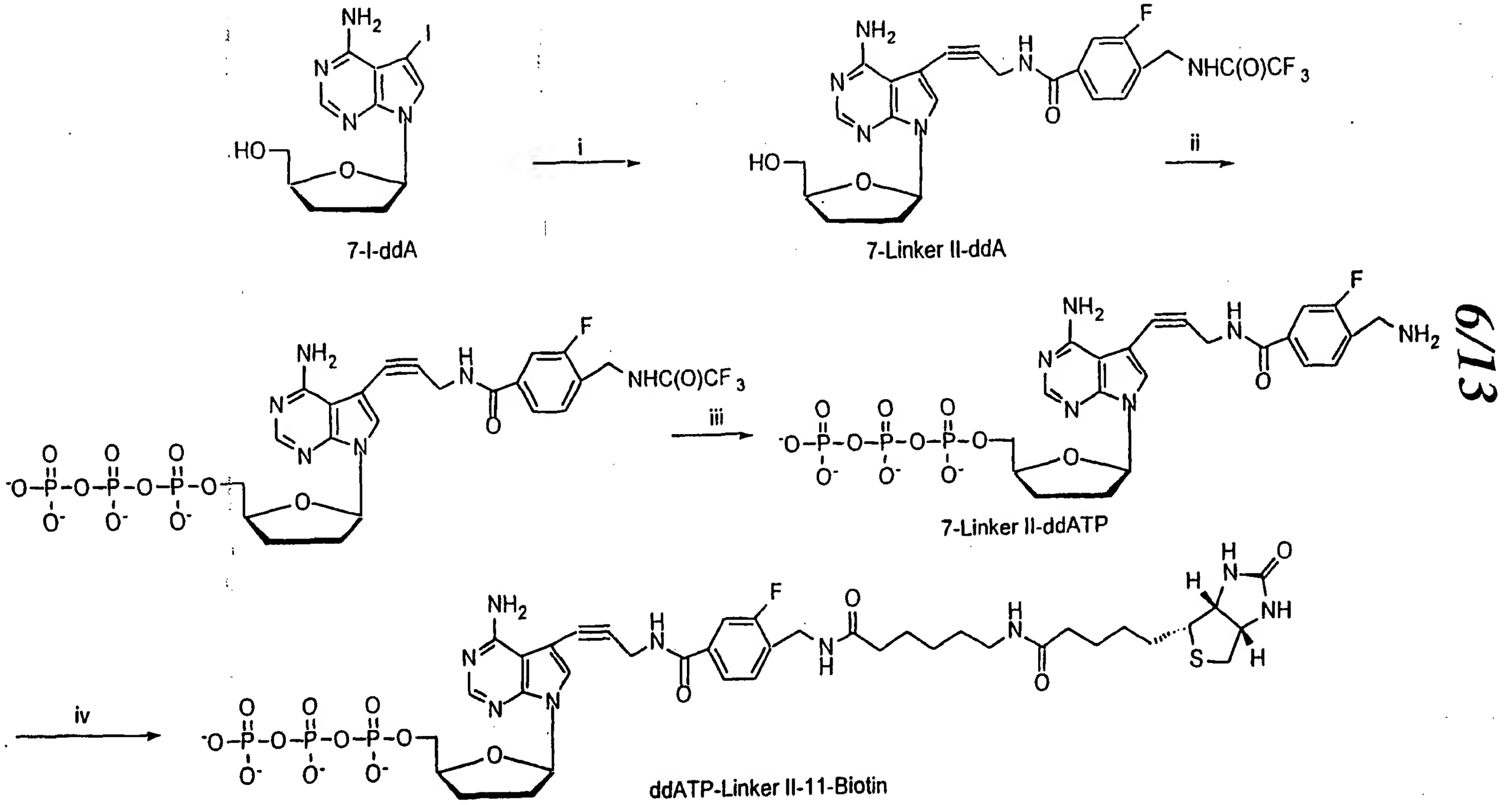
(Linker III, for ddG)



(Linker II, for ddA)

(i), $(CF_3CO)_2O$; (ii), Disuccinimidylcarbonate/diisopropylethylamine;
(iii), Propargyl amine

FIGURE 6



(i) Linker II, tetrakis(triphenylphosphine)palladium(0); (ii)$POCl_3$, Bn_4N^+pyrophosphate; (iii) NH_4OH; (iv) Sulfo-NHS-LC-Biotin

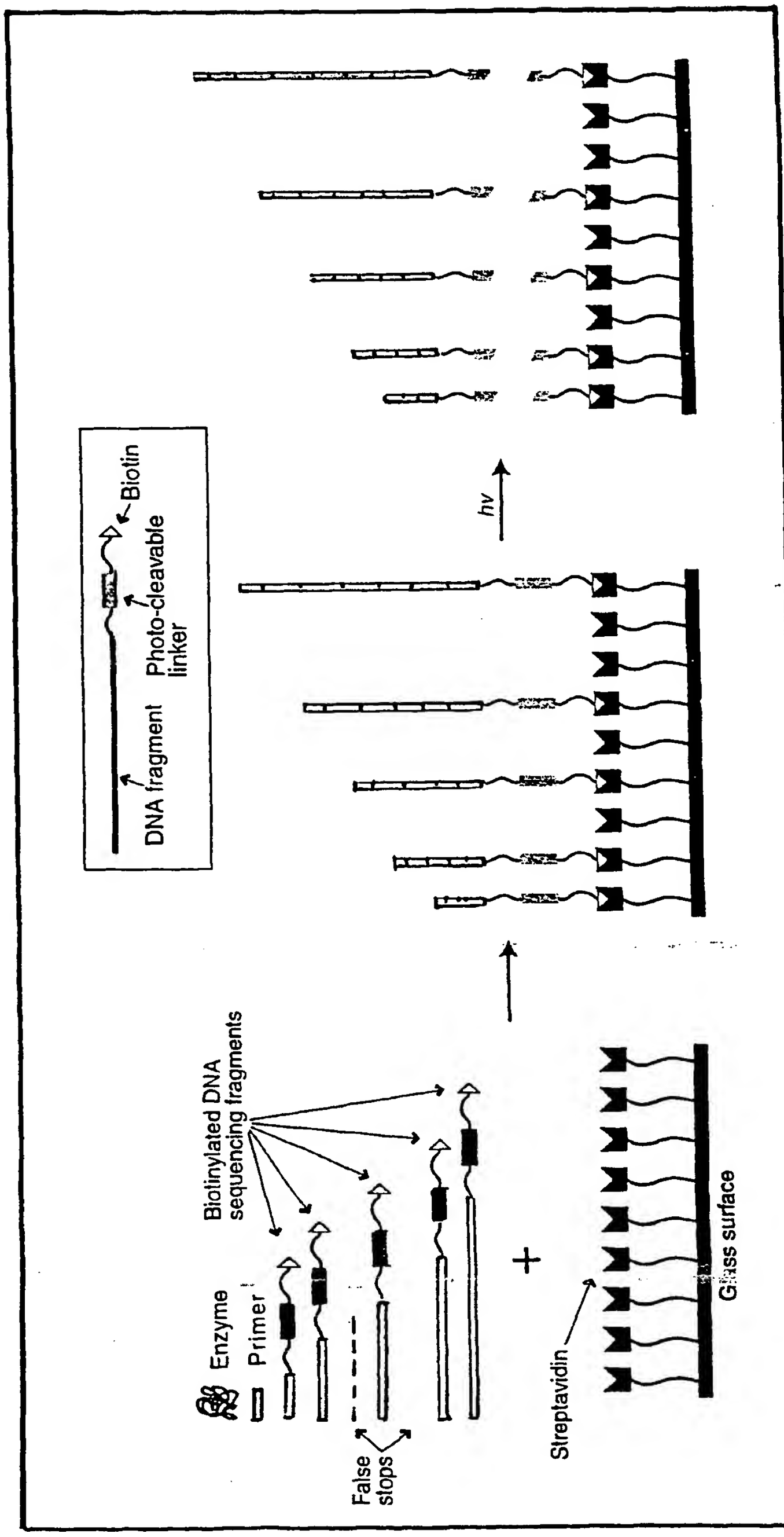

Biotin
Photo-cleavable linker
DNA fragment
hv
Biotinylated DNA sequencing fragments
Enzyme
Primer
False stops
Streptavidin
Glass surface

FIGURE 7

FIGURE 8

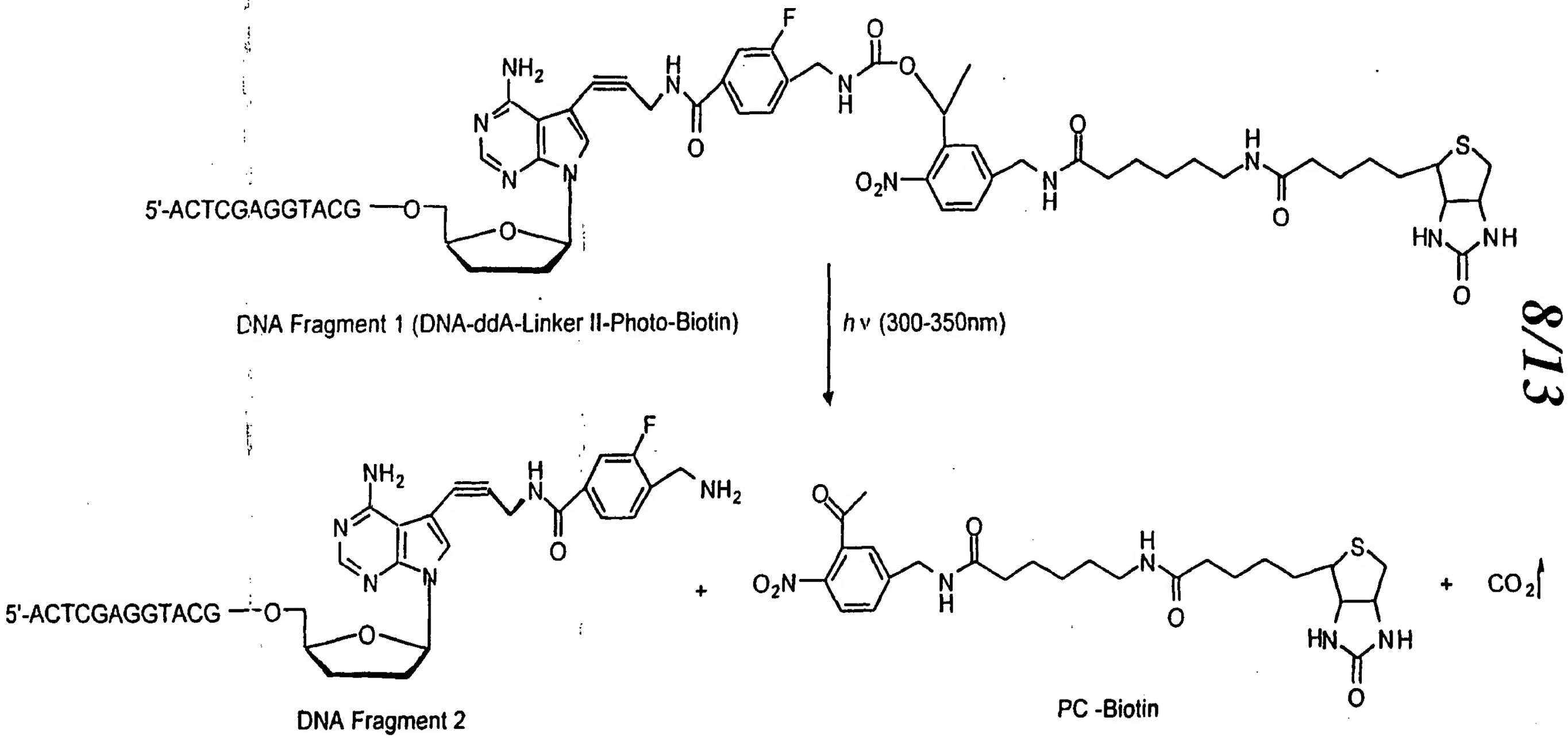

5'-ACTCGAGGTACG
DNA Fragment 1 (DNA-ddA-Linker II-Photo-Biotin)
hν (300-350nm)
5'-ACTCGAGGTACG
DNA Fragment 2
PC -Biotin
CO2

FIGURE 9

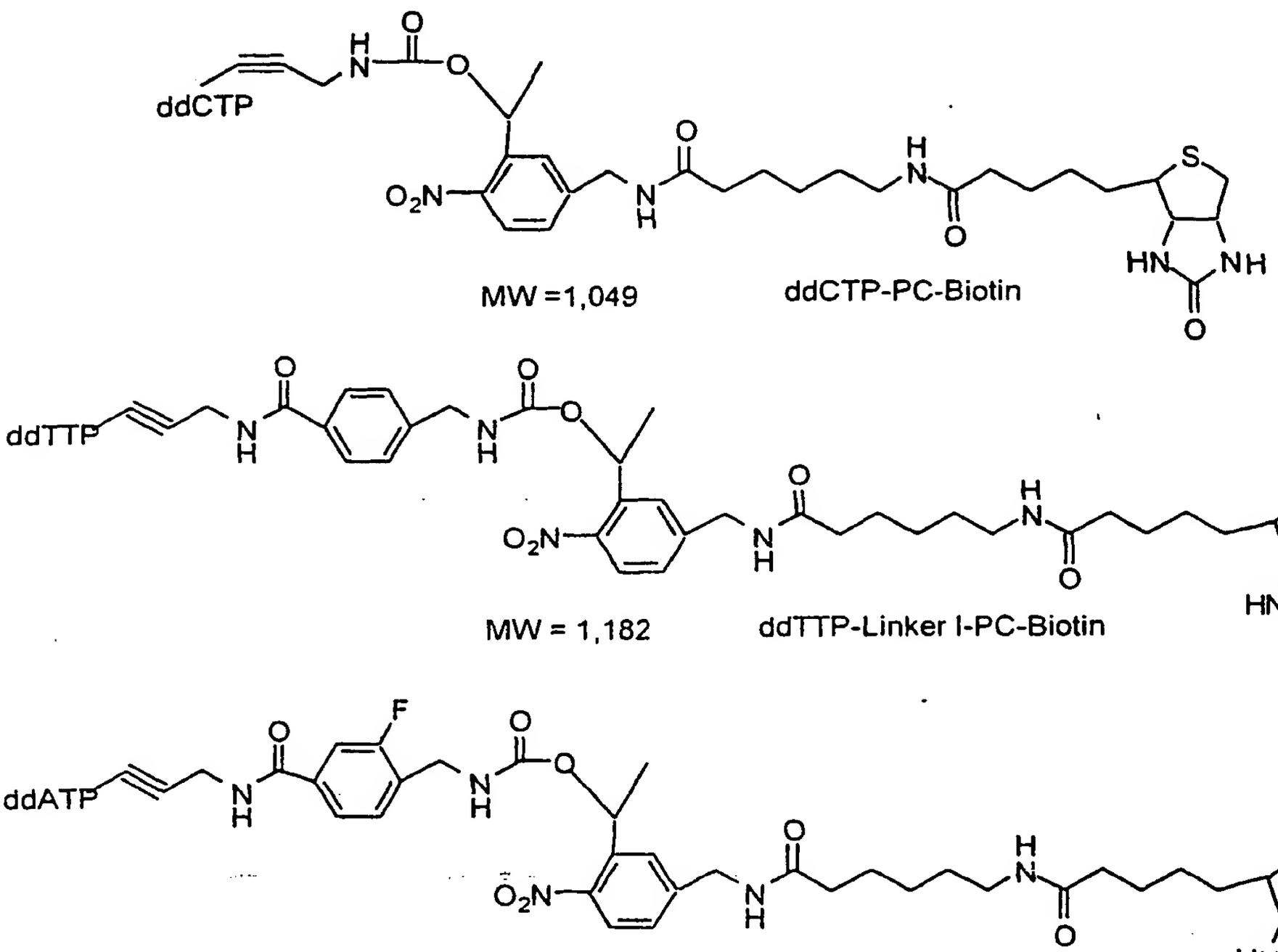
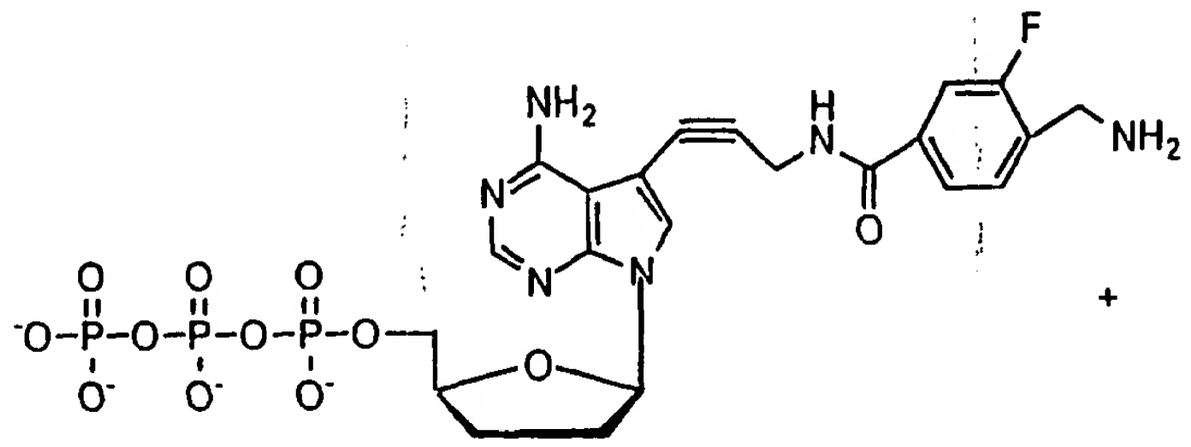
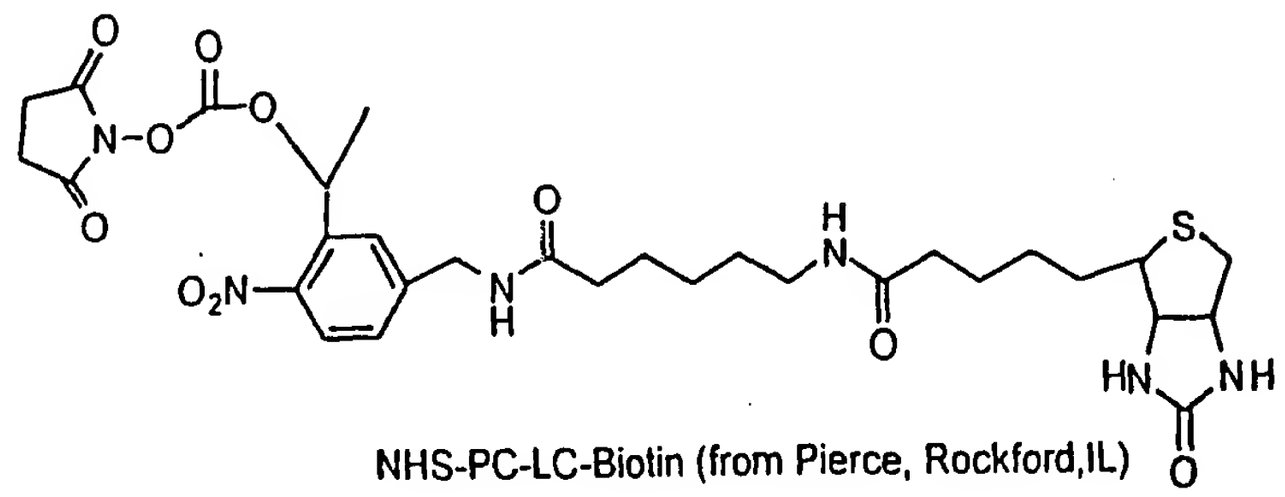

ddCTP
MW =1,049
ddCTP-PC-Biotin
ddTTP
MW = 1,182
ddTTP-Linker I-PC-Biotin
ddATP
MW = 1,222
ddATP-Linker II-PC-Biotin
ddGTP
MW = 1,257
ddGTP-Linker III-PC-Biotin

FIGURE 10

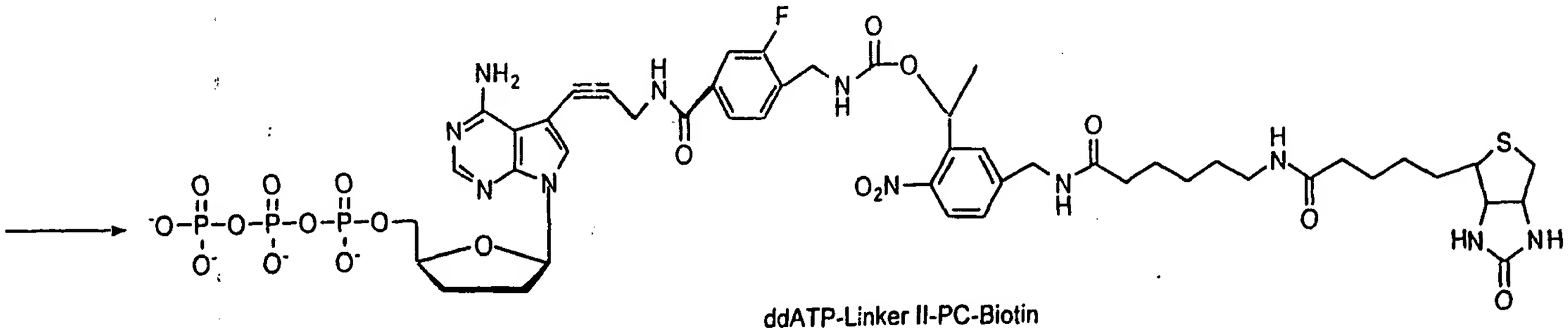


ddATP-Linker II

+



NHS-PC-LC-Biotin (from Pierce, Rockford,IL)



ddATP-Linker II-PC-Biotin

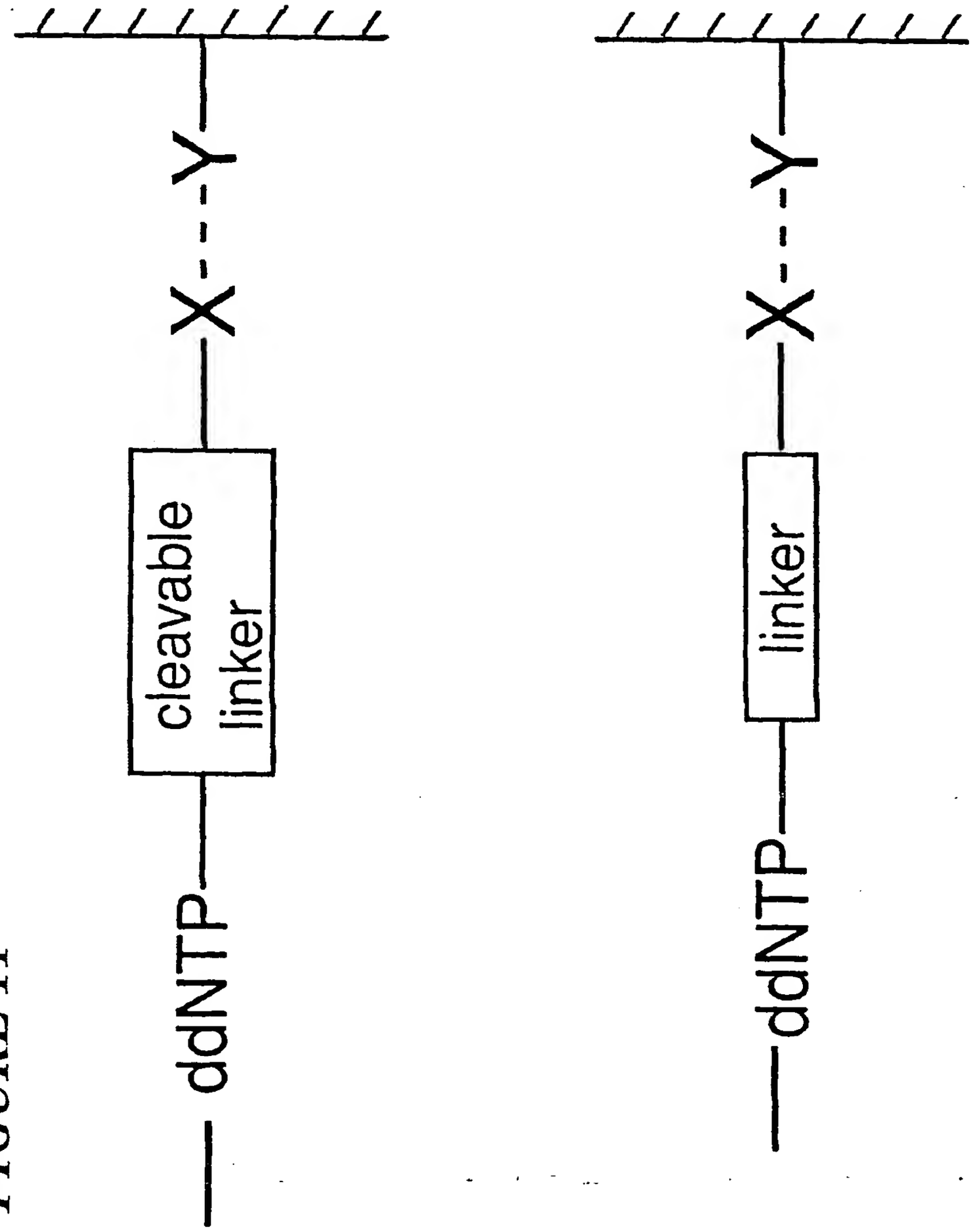


FIGURE 11

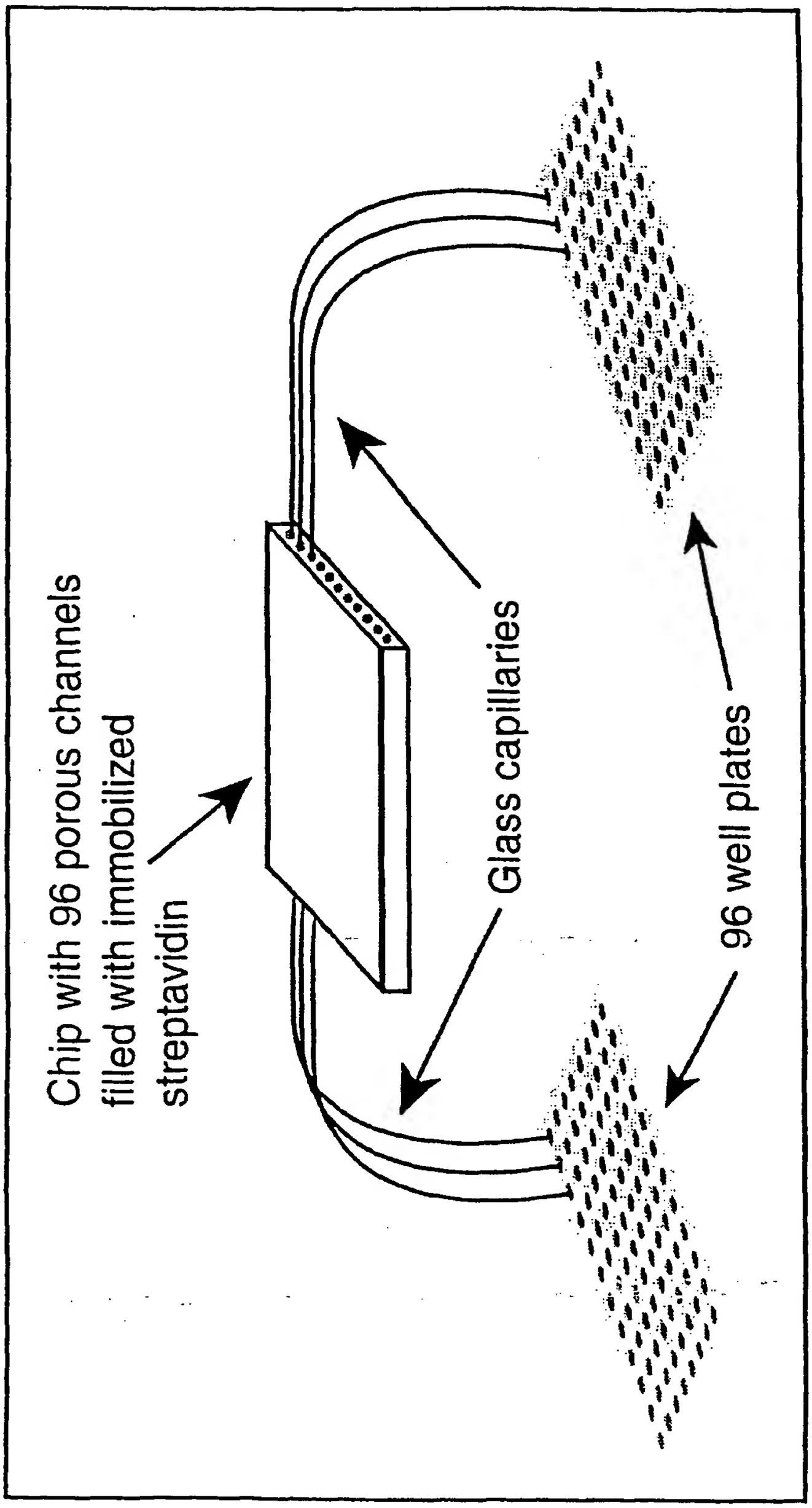
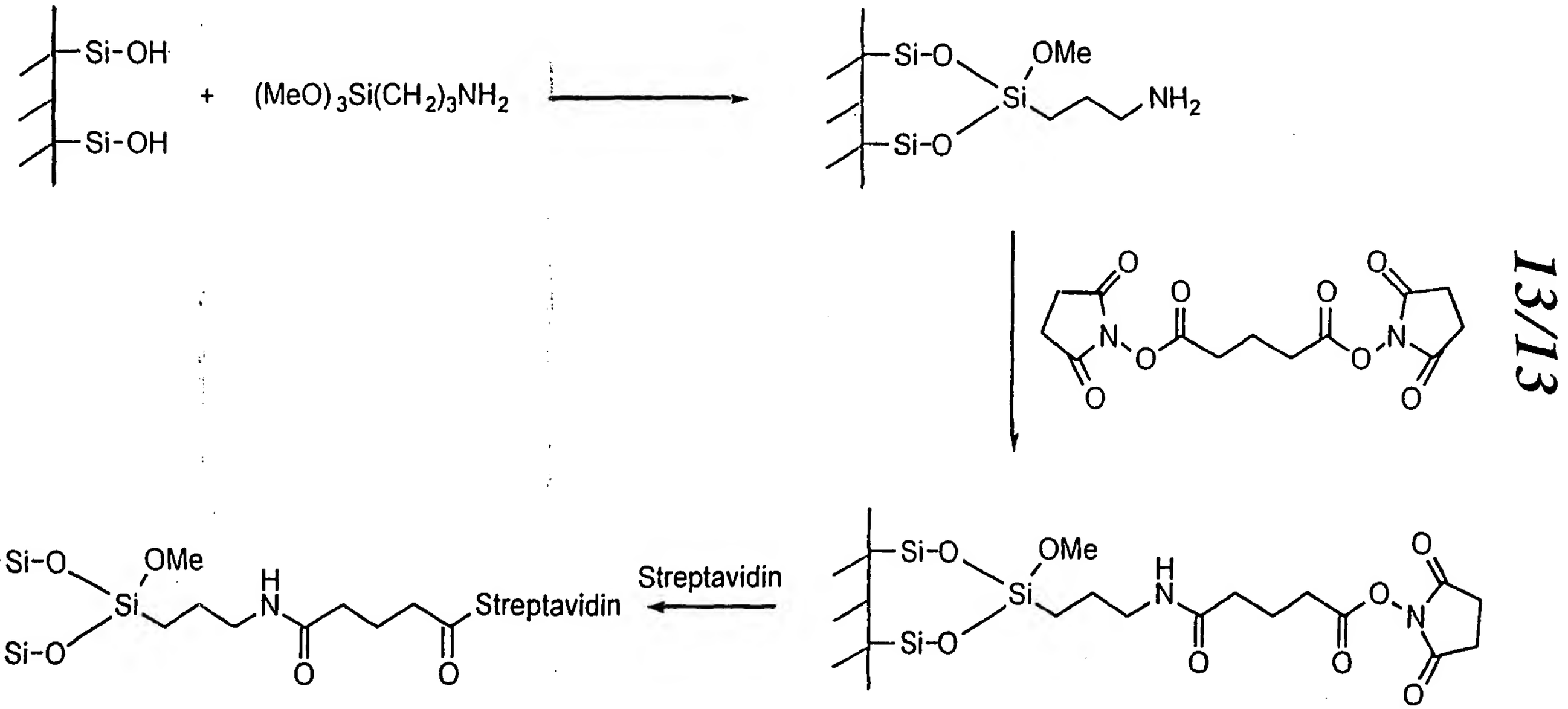

Chip with 96 porous channels
filled with immobilized
streptavidin
Glass capillaries
96 well plates

FIGURE 12

FIGURE 13


Si-OH
Si-OH
+
(MeO)3Si(CH2)3NH2
Si-O
OMe
Si
NH2
Si-O
O
N
O
O
O
O
O
O
N
O
Si-O
OMe
Si
H
N
Si-O
Streptavidin
Streptavidin
Si-O
OMe
Si
H
N
Si-O
O
O
O
O
N
O

SEQUENCE LISTING

<110> The Trustees of Columbia University in the City of New York

<120> High-Fidelity DNA Sequencing Using Solid Phase Capturable Dideoxynucleotides And Mass Spectrometry

<130> 0575/62948-PCT/JPW/ADM/BJA/AX

<160> 6

<170> PatentIn version 3.0

<210> 1

<211> 55

<212> DNA

<213> synthetic: template

<400> 1
acttttact gttcgatccc tgcatctcag agctcgctat tccgagctta cacgt 55

<210> 2

<211> 13

<212> DNA

<213> synthetic: primer

<400> 2
taaggctcga atg 13

<210> 3

<211> 61

<212> DNA

<213> synthetic: template

<400> 3
taaagctata ggtacagtat tagtaggacc tacacctgtc aacataatgg tccaggtcgt 60

g 61

<210> 4

<211> 13

<212> DNA

<213> synthetic: primer

<400> 4
ccaggtccag cac 13

<210> 5

<211> 129

<212> DNA

<213> synthetic: template

<400> 5
catgtgtaac agttcctgca tgggcggcat gaacccgagg cccatcctca ccatcatcac 60

actggaagac tccagtggta atctactggg acggaacagc tttgaggtgc atgtttgtgc 120

ctgtcctgg 129

<210> 6

<211> 15

<212> DNA

<213> synthetic: primer

<400> 6

ccaggacagg cacaa 15

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/09752

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :C12Q 1/68
US CL :435/6, 91.1; 536/23.1

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 91.1; 536/23.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WEST, STN, MEDLINE, BIOSIS, CAPLUS, EMBASE
search terms: dideoxynucleotides, incorporation, mass spectrometry

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 6,046,005 A (JU et al.) 04 April 2000, see entire document.	1-24, 33-58
Y	ARBO et al, Solid Phase Synthesis of Protected Peptides Using New Cobalt (III) Ammine Linkers INT. J. Peptide Protein Res. 1993. Vol. 42, pages 138-154, see entire document.	25-32
X,P	US 6,316,230 B1 (EGHOLM et al) 13 November 2001, see entire document.	1-58
Y,P	US 6,218,118 B1 (SAMPSON et al) 17 April 2001, see entire document.	1-58
Y	US 5,174,962 A (BRENNAN) 29 December 1992, see entire document.	1-58

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
28 MAY 2002

Date of mailing of the international search report
18 SEP 2002

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. (703) 305-3230

Authorized officer
Valerie Bell-Harris for
ARUN CHAKRABARTI
Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet) (July 1998)★